DOCTOR OF PHILOSOPHY

Identification and Characterisation of the Barley Row-Type Gene VRS3

Bull, Hazel Joanne

Award date:
2015

Identification and Characterisation of the Barley Row-Type Gene *VRS3*

Hazel Joanne Bull

A thesis submitted for the degree of

Doctor of Philosophy

The University of Dundee

September 2015

Declaration

I declare that I am the author of this thesis; and that, unless otherwise stated, all references cited have been consulted by myself. The research for which this thesis is a record has been carried out by myself and it has not been previously accepted for a higher degree.

This is to certify that Hazel Bull has carried out her research under my supervision and has fulfilled the ordinances and regulations of the University of Dundee, so that she is qualified to submit the following thesis in application for the degree of Doctor of Philosophy.

Dr Bill Thomas
The James Hutton Institute

Prof. Andrew J. Flavell
School of Life Sciences
University of Dundee

Prof. Robbie Waugh
The James Hutton Institute

Acknowledgements

I would like to thank my supervisors, Dr Bill Thomas, Professor Robbie Waugh and Professor Andy Flavell for their invaluable guidance and encouragement through all stages of this research and for their helpful critique whilst writing.

Throughout this project I have sought help and advice from a number of colleagues at The James Hutton Institute and James Hutton Limited, especially the Barley Genetics group, for which I am very grateful. In particular I would like to acknowledge Nicky Bonar, Allan Booth, Nicki Cook, Arnis Druka, Kelly Houston, Malcolm Macaulay and Gaynor McKenzie for sharing their molecular biology expertise; Craig Simpson for his help and advice with splicing and gene expression analyses; Richard Keith, Chris Warden and Derek Matthews for assistance with field operations; Jenny Morris, Clare Booth and Louise Donnelly at the JHI Genome Technology facility for their DNA sequencing services and RNA extraction advice; Christine Hackett for her help with statistical analyses and Philip Smith for his thorough proof-reading of this document.

This work was funded through a grant from the Mylnefield Trust and I would like to thank them and Dr.Nigel Kerby, former Director of Mylnefield Research Services for their support.

Lastly, I would like to express gratitude to my family and friends; in particular, my parents Penny and Will Bull and my brother David; also to Richard Denham and Jenny Christie; your advice and encouragement throughout has been greatly appreciated.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AFLP—Amplified fragment length polymorphism

AHDB—Agriculture and Horticulture Development Board

AMBA—American Malting Barley Association

BAC—Bacterial artificial chromosome

bp—base pair

BOPA—Barley Oligo Pooled Array

cM—Centimorgan

DNA—Deoxyribonucleic acid

FLcDNA—Full length complimentary deoxyribonucleic acid

GWAS—Genome-wide association analysis

GS—Growth Stage

Kbp—Kilobase pair

LD—Linkage Disequilibrium

LOD—Logarithm of the odds

NRQ—Normalised relative quantification

PCR—Polymerase chain reaction

pOPA—Pilot oligo pooled array

qRTPCR—Quantitative real-time PCR

SNP—Single nucleotide polymorphism

SSR—Simple sequence repeat

REML—Residual maximum likelihood

RFLP—Restriction fragment length polymorphism

RL—Recommended List

RNA— Ribonucleic acid

Abstract

Barley row-type describes the number of grains present at a node on the barley spike. Two forms exist amongst cultivated barley: two-rowed with only the central of three spikelets fertile producing a single grain at a node and six-rowed with all three spikelets fertile, producing three grain at a node. Twelve regions of the barley genome have been associated with the row-type character with specific genes identified at three loci, *VRS1*, *VRS4* and *INT-C*. Advancements in the understanding of the genetic control underpinning barley row-type enables the identification of potential mechanisms for improving yield and yield architecture within the cereals.

This study used genetic linkage mapping in segregating F_2 populations to refine the genetic location of the row-type locus, *VRS3*, to 16 candidate genes on barley chromosome 1H. Sequencing candidate loci in 32 *vrs3* induced mutant alleles identified *VRS3* to be a highly conserved JmjC histone demethylase, with two natural alleles within European cultivated barley.

VRS3 was further characterised as a potential means of improving grain uniformity within cultivated six-rowed barley, through phenotypic assessment of grain size in varying allele combinations of *VRS3*, *VRS1* and *INT-C*. The addition of six-rowed alleles at these loci was found to improve balance between central and lateral grain parameters, resulting in a more uniform grain sample.

Analysis of gene expression found *Vrs3* to be constitutively expressed across a diverse panel of barley tissues. Moreover, detailed study within the developing inflorescence suggests a role for *Vrs3* in the regulation of the row-type genes *VRS1* and *INT-C*.

.

1. General Introduction

1.1 *Global barley production*

With an average of 142 million tonnes produced globally (2005-2013) barley (*Hordeum vulgare* ssp. *vulgare*) lies fourth in the FAO list of global cereal crop production[1]. As a crop, barley can be grown in more marginal soils compared to other cereals[2], enabling it to be farmed in a wide range of environments from sub-Saharan Africa to Alaska and Norway (Figure 1).

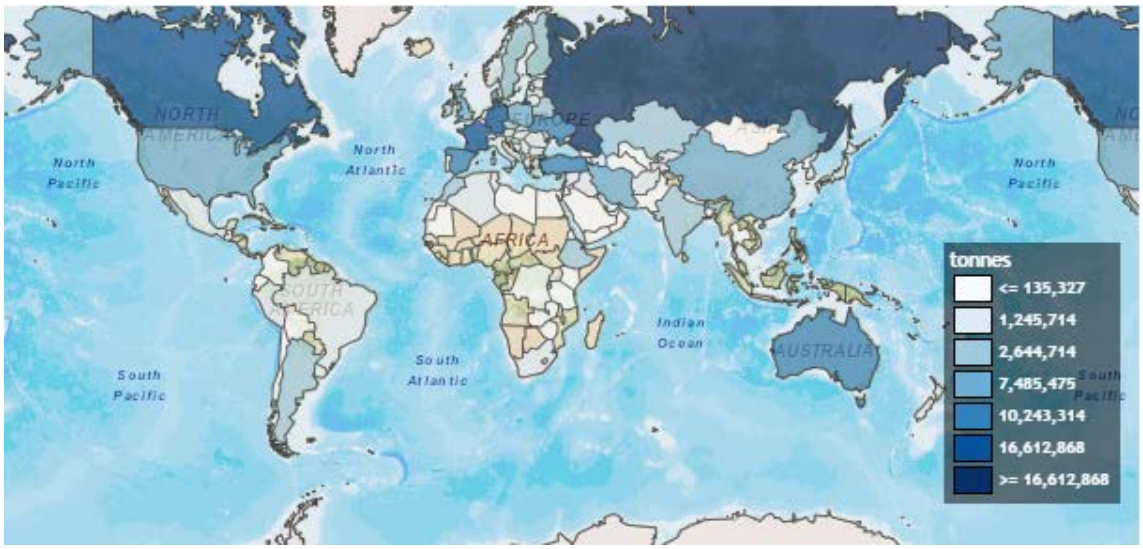


Figure 1: Average global barley production by country (2000-2013)
(*Adapted from FAOStat[1])*

Barley is grown primarily for its grain; either for post-harvest processing into malt, the primary ingredient in beer and whisky, or for use as an animal and human foodstuff. It is estimated that 85% of global barley production is used for feed[3] yet it is the premium malting market which is the major focus of commercial barley breeding programmes in North-west Europe.

The main aims of barley breeding are to improve yield and agronomic performance under variable environmental conditions whilst maintaining the strict quality requirements of end-users such as the malting, brewing and distilling industries.

Global malt production is approximately 18 Mt[4] and Europe has 42% of the global malting capacity and its exports generate €850,000,000 in revenue[5].

Quality requirements of the malting industry vary widely and are largely dependent on the downstream use of the malt[6]. Simplistically, the malting process is the controlled

germination of barley grain to maximise conversion of starch endosperm into fermentable sugars whilst minimising losses due to the growth of the rootlets and coleoptile of the developing barley seedling. The conversion of the starch into fermentable sugars, principally maltose, is known as modification.

One particular important malting quality criterion is grain size and uniformity. Grain size is measured as percentage screenings: the proportion of grain that passes through a sieve of set size, the size varying from country to country. In Europe the general specification for grain size is >90% of grain retained over a 2.5 mm sieve; in England and Wales this is slightly different with >94% over a 2.25 mm sieve[6]. Uneven grain size within a malting barley batch can result in uneven rates of uptake of water and uneven levels of modification[7,8].

Following modification and kilning the malted grain is milled to form grist. Non-uniform grain size can also impact on the uniformity of the milling process and the final grist[9]. A number of factors are known to influence barley grain size; these include stress at anthesis and grain-fill[8,10,11] and the morphology of the barley spike[12,13].

1.2 *Monocot inflorescence morphology*

1.2.1 *General distinctions*

Taxonomically barley (*Hordeum vulgare ssp vulgare*) is part of the Poaceae family of monocots, which also includes other cereal crops such as: *Triticum* (wheat), *Zea* (maize) and *Oryza* (rice). Despite being members of the same family, barley, wheat, maize and rice all share distinct morphologies particularly with respect to the inflorescence.

Maize is characterised by the production of a single stem per plant; it is also monoecious, with separate tassel and spike inflorescence structures (Figure 2b). The branched tassels are located at the apex of the stem, whilst the kernel producing spikes form above leaf axils. The maize morphology ideotype is in contrast to that of wheat, barley and rice which in addition to producing multiple tillers per plant are also all hermaphrodite. In the case of rice each tiller produces a highly branched inflorescence structure known as a panicle (Figure 2a). At the end of each branch a spikelet is formed, fertile florets within which lead to grain production.

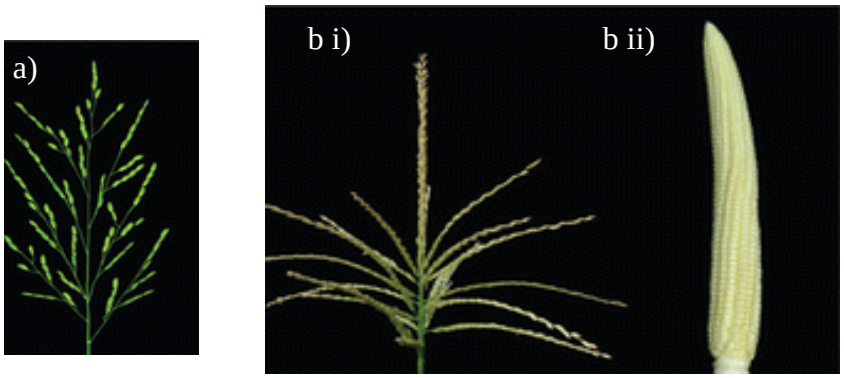


Figure 2: The inflorescence structures of a) rice and b) maize: i) male tassel ii) female spike
(adapted from Wang and Li, Ann. Rev. Plant Biol. (2008) 59: 253-279.)

Wheat and barley produce a single spike per tiller; these are less branched inflorescences with spikelets attached at nodes on alternate sides of a central rachis. One of the primary distinctions between wheat and barley spikes is the spikelet and floret structure. In wheat a single spikelet is present at each rachis node but each wheat spikelet comprises multiple grain producing florets (Figure 3). Conversely in barley, each spikelet is uni-floretted but generally up to three spikelets are present at a single rachis node (Figure 4-6).

Figure 3: The structure of the wheat spike
a) An entire wheat spike; b) spikelets attached at nodes on alternate sides of the central rachis; c) a single spikelet comprised of multiple florets; d) a wheat spikelet with a single floret open exposing the anthers. Scale bars represent 1cm.

Differences in plant architecture are regulated through meristems, collections of undifferentiated plant cells. The regulation of the differentiation of the cells within the meristem results in distinct morphologies. All above-ground regions of the plant originate from the shoot apical meristem (SAM) which is formed during embryogenesis[14]; SAM can undergo further post-embryonic differentiation to form axillary meristems[15], and it is from these axillary meristems that the lateral branches of the plant, including tillers, arise. The uniculm nature of cultivated maize is due to suppression of outgrowth at the axillary meristems[16].

The inflorescence is formed from the switch of the SAM from vegetative to reproductive growth. Following the transition, the SAM forms spikelet meristems, either with or without the formation of branched meristems; the spikelet meristem in turn forms floral meristems which differentiate to form the floral organs of the plant.

In rice the floral meristems produce primary and secondary branch meristems prior to spikelet formation resulting in the panicle architecture. In contrast, the spikes of wheat and barley do not produce branch meristems prior to spikelet meristem formation, resulting in the spikelet attached directly to the central rachis. The maize tassel is intermediate between the two, with multiple primary branches formed from the SAM; subsequently very short spikelet pair meristems branch to form spikelet meristems, one of which aborts resulting in a single floret per spikelet pair meristem.

The genetic and hormonal controls of branching in rice and maize have been studied in some detail[17-20] but the controls in wheat and barley are still less well understood.

1.2.2 *The barley inflorescence*

Barley spikelets are constructed of two leafy structures, the lemma and palea, which enclose the floral organs; each spikelet is subtended by a pair of glumes (Figure 4). A characteristic feature of grain producing spikelets of most cultivated barley is the extension of the lemma tip into a narrow awn. However, the awn is not necessary for grain production, with awnless two-rowed and six-rowed cultivars bred particularly in North America.

At the base of the adaxial side of the spikelet is a narrow rod-shaped projection called the rachilla. Opening of the barley floret at anthesis is controlled by swelling of the lodicules; however, the lodicules of most cultivated barley are reduced resulting in the floret remaining closed at anthesis and therefore undergoing self-pollination[21].

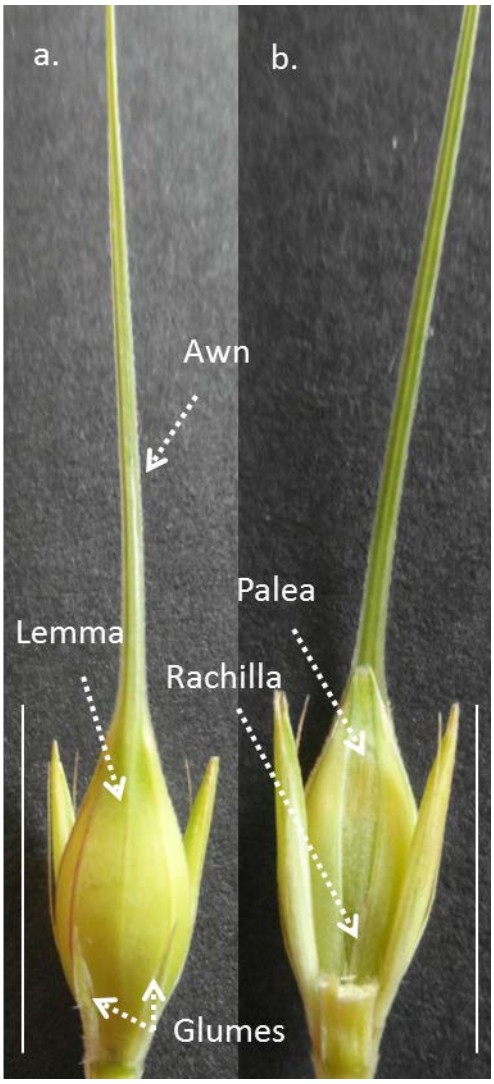


Figure 4: The barley spikelet

a)The abaxial view with lemma, awn and glumes visible; b)the adaxial view with palea and awn visible, whilst not visible the approximate position of the rachilla is also shown. Scale bar represents 1 cm.

Barley inflorescence development initiates from the shoot apical meristem relatively early on in plant development, at approximately the 2—3 leaf stage (Zadoks growth stage 12—13[22]). The changes in morphology during this time have been studied in detail by Kirby *et al*.[23]. Spikelet differentiation occurs acropetally, with new spikelet ridges continually differentiated from the spike apex until maximum differentiation at the awn primordium stage. Despite the gradation in differentiation, it is the spikelet primordia in the central region of the inflorescence that reach full differentiation first, with spikelets at the base of the spike slower and some spikelet primordia at the tip of the inflorescence aborted prior to full differentiation.

Initially the dome shaped meristem differentiates to form multiple pairs of ridges (double ridge stage); the upper of each ridge pair proceeds to form the spikelet whilst

the lower, the leaf ridge, forms leaf structures such as the flag leaf and collar in the lower portion of spike but becomes more indistinct towards the spike apex.

Subsequently the spikelet ridge differentiates into three mounds, with the central spikelet meristem subtended by two lateral spikelet meristem (triple mound stage). Inflorescence development then proceeds with the initiation of development of glumes (glume primordium), lemma (lemma primordium) stamen (stamen primordium) and finally awns (awn primordium).

Unique to barley, variable development and fertility of the spikelets at a rachis node results in very distinct inflorescence morphologies within the species. This distinction in morphology is used as a sub-classification of barley cultivars, classifying them as either two-rowed or six-rowed. The distinction between two-rowed and six-rowed spikes occurs after the triple mound stage of inflorescence development[23],

1.2.2.1 *Two-rowed*

Two-rowed barley spikes are characterised by only the central of the three spikelets at a rachis node being awned, fertile and hence producing grain. The neighbouring, lateral spikelets, are reduced in size, with rounded lemma and no awn (Figure 5).

Figure 5: The two-rowed barley spike
a) the entire spike; b) spike with the lower spikelets removed exposing the alternate rachis nodes to which the spikelet triplets are attached; c) a spikelet triplet with the central spikelet fertile, awned and filled with grain, the neighbouring infertile lateral spikelets are reduced in size and with rounded lemma. Scale bars are equal to 1 cm.

1.2.2.2 *Six-rowed*

By contrast, in six-rowed barley spikes all three spikelets at a rachis node are enlarged, fertile, awned and grain producing. However, the grain produced by the lateral spikelets are smaller compared to the central spikelets[24].

Figure 6: The six-rowed barley spike
a) the entire spike; b) spike with the lower spikelets removed exposing the alternate rachis nodes to which the spikelet triplets are attached; c) a spikelet triplet with both the central and lateral spikelets fertile, awned and filled with grain. Scale bar represents 1cm.

1.2.3 ***Tillering***

Barley tillers extend from tiller buds formed in the axils of leaves. The first stem produced by the barley plant is identified as the main stem. In the leaf axils of the main

stem primary tillers are formed in a sequential order with primary tiller 1 (T1) extending from the axil of leaf 1, primary tiller 2 (T2) extending from the axil of leaf 2 etc. The number of primary tillers produced is known to vary depending on a number of environmental factors including sowing density, water and nitrogen availability but is typically 3-4 per plant[25,26]. In addition to the primary tillers, under some environmental conditions, typically those not limited by water or nitrogen availability and low sowing density, secondary tillers and tertiary tillers can arise from the leaf axils of the primary and secondary tillers respectively in the same sequential manner[25]. Typically spike yield from secondary and tertiary tillers is reduced compared to that of spikes originating from primary tillers or the main-stem [25]. Tiller number can vary greatly during the growing cycle of the plant, with tillers both produced and killed; therefore not all tillers produced contribute to overall yield of the plant. The reasons for this remain unclear but it is thought it may be a stress response of the plant, with tiller numbers reduced at times of environmental stress such as reduced nutrient availability[25,27].

1.2.4 ***Barley gene nomenclature***

Standard barley genetic loci nomenclature abbreviates loci to an italicised three character symbol representing the phenotype; capitalisation of the first letter represents a dominant character and lower case a recessive character e.g. *Int* for two-rowed spike and *int* for intermedium spike. Non-allelic loci are further distinguished by a hyphen and letter or a number e.g. *intermedium-a* is abbreviated to *int-a* and *intermedium-b* to *int-b*. Independent alleles at a locus are identified by a full-stop and a non-italicised letter or number following the locus symbols e.g. *int-a*.1 is allele 1 of the *intermedium-a* locus and *int-a*.2 is allele 2 of the *intermedium-a* locus.[28]In some cases an ‘H’ is included in gene nomenclature following the locus symbol to distinguish the barley locus from other *Triticeae* e.g. *VrnH2* or alternatively ‘Hv’ prior to the gene symbol e.g. *HvTb1*.

1.3 ***Barley growth habit***

A second key sub-division of the cultivated barley crop is that of growth-habit, either spring or winter. Barleys with a winter growth-habit have a vernalisation requirement of approximately 6 weeks below 4^{o}C to promote inflorescence development whereas barley with a spring growth habit has no chill requirement. The requirement for vernalisation is controlled through allelic variation and epistatic interaction at the *VRNH1* and *VRNH2* loci. In winter growth habits the *VrnH2* locus is present but absent

in spring types[29]. The level of vernalisation requirement is believed to be controlled through intron length variation at the *VRNH1* locus[29,30].

1.4 *Row-type and current barley cultivars*

In the U.K. growers are advised on the best barley cultivars to grow through the AHDB (Agriculture and Horticulture Development Board) Recommended List. Cultivars are divided into winter and spring growth habit and suitability for the respective markets: malt - distilling, malt- brewing and the animal feed market. Currently in the U.K. six-rowed barleys are only recommended for the winter animal feed market, with 4 of the 19 winter cultivars on the 2015/16 AHDB Recommended List six-rowed; all of the remaining 15 are two-rowed, with 5 recommended for malting for the brewing sector and 11 for the animal feed market. By contrast in France, six-rowed barleys are accepted for the premium malting market; currently 5 of the 7 winter varieties listed as preferred for malting in 2016 by the Malteurs De France are of the six-rowed type. In North America of the 25 varieties recommended for malting by the AMBA (American Malting Barley Association Inc.) in 2015, 7 are of the six-rowed spring type.

It is not clear where the preference for two-rowed barley by UK maltsters stems from; in his 1947 book[31], Beaven suggests six-rowed barleys fell out of favour with UK maltsters with the introduction of the malt tax in the late 17th century. Maltsters were taxed based on the volume of soaked grain; the smaller grain associated with six-rowed barley led to larger volumes of soaked grain compared to two-rowed types and therefore higher taxes.

Following the taxes repeal in the latter part of the 19th century, UK maltsters imported large volumes of Californian six-rowed barley in addition to the locally grown two-rowed varieties such as Chevallier and Spratt Archer. At this time, six-rowed barleys in England were only sown as a source of forage for animal feed. However, in the Scottish highlands, the traditional six-rowed land-race Scots Bere was grown both for human food consumption and the distilling industry, where it was subjected to lower malt tax due to its lower alcohol yield compared to other barley[32].

Indeed in the 1920's Beaven recognised a gap in the market for a UK grown six-rowed winter barley which could replace the six-rowed Californian types should supplies become scarce[31]. Beaven developed the variety F112, a single plant selection from a sample of grain purchased from the Dayton Blue Brewing company[33]. In large scale malting and brewing tests F112 was found to have an extended dormancy period but

once broken it was found to be a satisfactory substitute for imported American six-rowed types[34]. Although found to be suitable for malting it was not sold commercially primarily due to its susceptibility to the fungus *Cladosporium*.

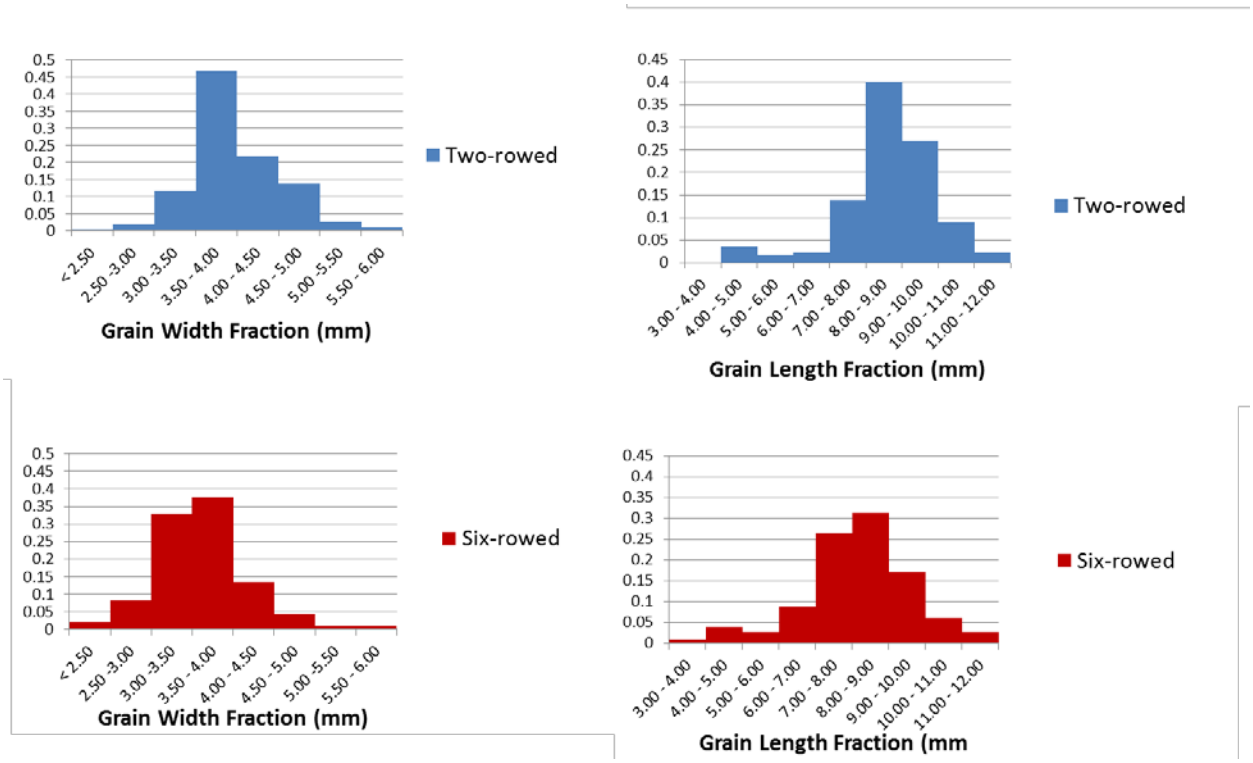


Figure 7: Comparison of the grain width and grain length size distributions from samples of grain of two-rowed (blue) and six-rowed (red) barley cultivars grown in Scottish field conditions in 2014.
On all graphs the y axis represents the proportion of sample in each size fraction.

The more variable grain size distribution associated with six-rowed barleys compared to two-rowed barleys (Figure 7) may also influence the selection of two-rowed barleys over six-rowed barleys for malting.

1.5 *Agronomic advantages of six-rowed barley*

Barley yield is influenced not only by the size and weight of individual grain but other factors including the number of fertile spikes produced per unit area and the numbers of grain per spike.

Genetic variability with respect to spike production has been identified; in particular, two-rowed cultivars in general produce a greater number of tillers per plant than six-rowed cultivars. Therefore despite producing three times as many grains per rachis node, under UK conditions, the difference in yield between current recommended six-rowed and two-rowed barley cultivars is negligible[35]. This is likely due to a combination of increased tiller number per plant and individual grain weight.

However, studies of barley over multiple environments suggest that the number of spikes per unit area is highly environmentally sensitive with low levels of heritability across environments[36,37]. In environments characterised by short growing seasons either through cold temperatures delaying sowing or the possibility of the onset of drought during the growth cycle, barley with rapid growth rate and an early spike emergence is preferential, to enable anthesis and grain-fill to be completed and yield potential maximised[38-40] . However, this model of plant development is also associated with a reduced number of tillers per plant[37], therefore under these conditions an increased number of grains per node would be favourable.

This is illustrated by the preference for the growth of spring six-rowed barley in the northern latitudes of Norway and Finland. In these environments the growing season is characteristically short, with prolonged cold delaying sowing, therefore six-rowed cultivars with fewer tillers but early flowering are preferred[39]. Harlan also noted a definite switch to the growth of six-rowed barleys in arid regions of Iran where irrigation was essential compared to the growth of two-rowed barleys in the cooler Zagros mountains[41].

Climate change predictions suggest an increase in average temperature during the growing season and an increase in extremes of weather[42]. A recent study into the response of barley cultivars to changes in climate found an increase in temperature in the weeks immediately pre and post anthesis dramatically reduced yield as did higher temperatures during the vegetative growth phase 3-4 weeks after sowing[43].

Whilst extended drought is not currently a problem in the UK, temporal drought stress does occur e.g. crop year 2011. If climate change predictions prove correct then temporal drought could occur more frequently and therefore six-rowed barleys could prove increasingly advantageous over their two-rowed counterparts with their increased grain per spike but fewer tillers per plant. However to fulfil malting quality requirements the uniformity between the central and lateral grain size would require improvement.

1.6 *The wild progenitors of barley and row-type*

The wild barley species *Hordeum vulgare* ssp. *spontaneum* (*H. spontaneum*) has long been recognised as the progenitor of cultivated two-rowed barley. Spikes of *H. spontaneum* are characterised by a two-rowed, brittle-rachis, spike phenotype, thus when the spike matures, spikelets disarticulate from the rachis and grain is dispersed.

The two-rowed phenotype is thought advantageous as the reduced lateral spikelets form an arrow shape to help project grain into the ground. The selection, intentional or otherwise, of a non-brittle rachis phenotype, with mature grain retained on the spike, is believed to be the major event underlying the domestication of barley[44-46].

The origins and domestication of six-rowed cultivars has been the subject of some debate, primarily due to its distinct morphology from *H. spontaneum* and two-rowed barley. Under Linnaean classification two-rowed and six-rowed barley were classified as separate species *H. distichum* and *H. hexastichon* respectively.

In his early investigations into the origins of cultivated plants[47], Candolle concluded that two rowed barley most likely originated in 'Western temperate Asia', however he found no evidence of a six-rowed barley progenitor species growing wild. He proposed two hypotheses, either six-rowed barley originated from the same wild ancestor as two-rowed barley or alternatively the wild-ancestor of six-rowed barley is extinct, although he noted there was no archaeological evidence to support the latter hypothesis[47].

In the 1930's evidence of six-rowed barley with a brittle rachis phenotype, was identified in grain collected from Tibet. This form of barley was classified as *H. agriocrithon*[48]. However, it was later concluded that no forms had been found growing truly wild but at margins with cultivated forms. Therefore the grain was most likely the product of a post-domestication cross-hybridisation event between cultivated six-rowed and *H. spontaneum* rather than a true wild-species[49-51].

Cross-pollination between two-rowed and six-rowed cultivars found them to be interfertile with subsequent generations both two-rowed and six-rowed, with six-rowed recessive to two-rowed[52]. It was therefore concluded that six-rowed barley arose from two-rowed barley following a mutational event post-domestication[41].

1.7 *Global spread of barley*

The primary region of barley domestication is recognised as the Fertile Crescent, a region of land which extends along the Euphrates and Tigris rivers, from the Persian Gulf to the Mediterranean and into the lower Nile delta. Here, diverse populations of *H. spontaneum* are found growing wild.

Whilst *H. spontaneum* has been identified in archaeological remains dating back 22,000 years, the non-brittle rachis two-rowed phenotype is only thought to have arisen around 8000 years BC. From the Fertile Crescent, *H. vulgare* is believed to have spread west

via the Mediterranean into Central and Northern Europe by 3000 BC[41]. Despite originating from two-rowed barley, archaeological evidence suggested that by 4000 BC, six-rowed barley was the predominant form of barley under cultivation in the Fertile Crescent[41].

Later introduction of barley into America occurred via multiple means, with the Spanish conquest responsible for the introduction of six-rowed barley into Central America and Southern USA in the 18th century and British and Dutch settlers introduced two-rowed and six-rowed barley respectively into Northern USA during the 1600's[53]. The introduction of barley into Australia is also thought to have followed British colonisation[53].

In addition to the well-documented spread from the Fertile Crescent, recent genetic analysis has suggested further barley domestication centres in Morocco, Ethiopia and Tibet[54,55].

H. spontaneum, is a highly genetically diverse species following multiple rounds of recombination and spontaneous mutations since evolution from its common ancestor and prior to its domestication. Whilst it is not clear precisely how many independent domestication events occurred, domestication of *H. spontaneum* to form *H. vulgare* occurred in a relatively small proportion of the total *H. spontaneum* germplasm resulting in a substantial loss of genetic diversity immediately post-domestication, a so called domestication bottleneck. As barley spread post-domestication, its diversity increased through highly localised selection of plants which performed best in the local agronomic conditions.

These locally adapted landraces form the basis of the pedigrees underlying current barley breeding. With respect to the breeding of six-rowed cultivars, many of the six-rowed German cultivars can be traced back to crosses between the landrace selections Ekkendorf-Mammut and Groninger[56], whilst French six-rowed pedigrees trace back to the landrace selection Hatif d'Grignon[56].

1.8 *Barley as a model genetic organism*

Since the early 1900's barley has been used as a model monocot cereal species. Its diploid nature makes the genetics underlying traits relatively simple to follow compared to those of higher ploidy cereals such as hexaploid wheat (*Triticum aestivum*). Although naturally inbreeding, cross hybridisation between plants is simple to manipulate by the

emasculation of spikelets and manual addition of pollen. Subsequent inbreeding enables traits to be considered largely fixed within a population after approximately 6 generations, with average heterozygosity at any one locus approximately 3%.

The total barley genome spans an estimated 5.1 Gigabases (Gb) over 7 chromosomes[57]; its relatively large size when compared to other model plant species such as rice (430Mb)[58] and Arabidopsis (135Mb)[59] has precluded detailed genome sequencing until relatively recently.

1.8.1 ***Molecular markers***

Over the last hundred years, the means and accuracy of ascertainment of trait linkage within the barley genome has evolved remarkably. Early initial associations relied on the analysis of co-inheritance ratios between morphological phenotypes following cross-hybridisation and these were relatively rare outside special mutant collections.

As technologies to detect variation at the protein and then the DNA sequence level became available, the ability to associate phenotypic variation with these characters increased. Initially these were protein based in the form of isozymes[60] but were subsequently replaced with DNA based marker systems principally RFLP (Restriction Fragment Length Polymorphism)[61], AFLP (Amplified Fragment Length Polymorphism) and SSR (simple sequence repeats)[62] for which barley was well characterised. SNP (single nucleotide polymorphism) are now the most extensively used marker system in barley with 4.3 million SNP anchored to the high density barley POPSEQ map[63].

In parallel with the development in molecular markers, the development of semi-automated high-throughput genotyping technologies, either genotyping at a single SNP over thousands of cultivars, e.g. KASP marker system (LGC Genomics), which uses competitive allele specific PCR primers labelled with tail sequences complementary to the differential fluorescence signals FAM and HEX[64], or a single genotype with 9000+ SNP in a single reaction e.g. Barley Illumina iSelect SNP chip[65,66] has made linking genotypes and phenotypes more accurate and more cost-effective.

With the improvement in molecular markers, the detail of barley reference genetic marker maps has also improved. Early SSR and RFLP reference marker maps comprised approximately 30—40 recombination events per chromosome[61,62] but the most recent SNP consensus maps now contain from 580—1200 SNP recombinations per chromosome[66,67]. The high frequency of SNP throughout the genome has enabled

the resolution to which traits can be mapped within the barley genome to be improved dramatically.

1.8.2 *Synteny between cereals*

Conserved gene order across genomes of members of the grass family was first identified through the cross hybridisation of RFLP probes from wheat. A crop circle model was used to show the alignment and overlap of conserved genome regions[68]. This model has been gradually developed and improved with the improvements in genotyping and sequencing. Figure 8 illustrates the conserved gene order between rice and barley.

As alluded to earlier the large size of the barley genome has precluded detailed characterisation of the genome and local gene order to the same extent as smaller genome species such as rice. Although techniques such as genome walking were available, they were relatively inefficient in barley with a 100Kb BAC typically containing only 2—3 genes[69]. The presence of large numbers of repetitive elements within the barley genome also hindered the formation of a minimum tiling pathway[69,70].

The fact that barley has strong synteny with rice has enabled regions to be delimited with molecular markers in the barley genome and subsequent identification of the conserved region within rice can be used to identify candidate genes. This technique has been used successfully in the identification of a number of barley genes[71,72].

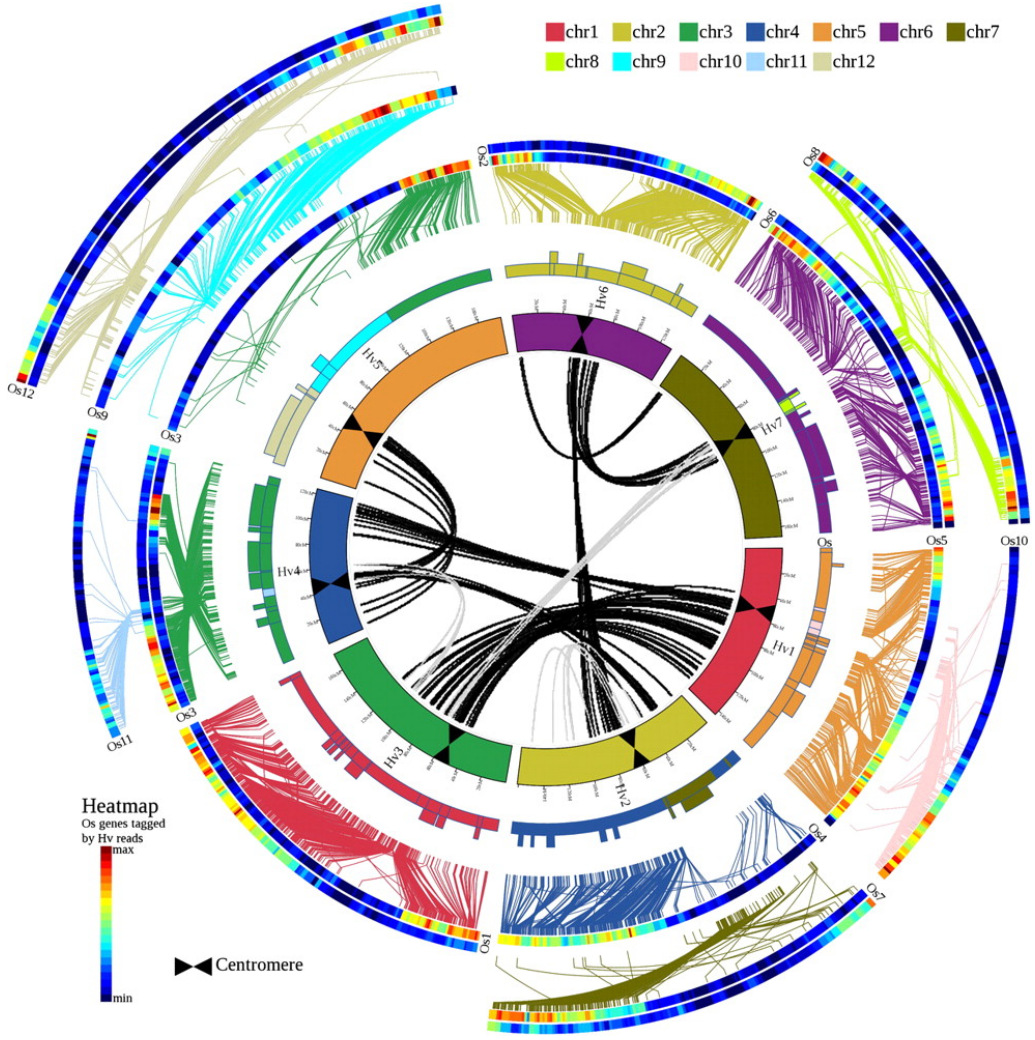


Figure 8: The conserved gene order between regions of the rice and barley genomes *(Reproduced from Klaus F.X. Mayer et al[73]. Plant Cell 2011;23: 1249-1263) The two inner-most circles represent the seven chromosomes of barley, and the outer circles the 12 chromosomes of rice. The colours of the barley chromosomes depicted in the outer of the two barley chromosome rings represent the region of the respective rice chromosome with which barley chromosome shares conserved gene order.*

1.8.3 ***Barley sequence resources***

1.8.3.1 *Morex genomic sequence*

The draft reference barley genomic sequence from the North American elite six-rowed cultivar Morex was publically released in 2012[57]. An outline physical genome scaffold was constructed through the sequential alignment of BAC (Bacterial artificial chromosome) sequences[74] onto which BAC genomic DNA sequence and Morex whole genome sequence contigs were anchored[57]. Sequence in the transcribed regions of the genome was enriched by the anchoring of Full length cDNA sequence from the cultivar Haruna Nijo[75] and RNA sequence data originating from 8 different Morex growth stages[57].

Through comparison with other reference plant genome sequences: rice, *Sorghum*, *Brachypodium* and Arabidopsis it is estimated that the current genome sequence represents approximately 80% of the entire barley genome and approximately 90% of the gene containing space[57]. Mapping of RNAseq reads against the genome scaffold was used to identify transcribed portions of the barley genome. The sequences of 26,159 genes were annotated as high confidence due to a high level of sequence homology with at least one reference sequence: rice, *Sorghum*, *Brachypodium* or *Arabidopsis* and evidence of gene clustering, that is genes from the same family found within the same region of chromosome, likely due to gene duplication events. A further 53,220 genes were described as low-confidence as despite transcripts mapping to the genome scaffold no evidence of homology or gene family clustering could be found to support their annotation[57]. Whilst recognised as not a fully annotated complete genome sequence, the Morex reference sequence represents a huge step forward in barley research.

1.8.3.2 *Full length cDNA libraries*

In barley, 12 full length cDNA (FlcDNA) libraries have been created and resequenced from the two-rowed Japanese malting barley cultivar Haruna Nijo. To maximise the opportunity of capturing the majority of expressed barley genes, the RNA for the creation of each library was isolated either from different tissue types or from plants grown under different environmental stresses. The 12 libraries were re-sequenced using a primer-walking method[76]; subsequent sequence alignment resulted in the identification of 22,651 non redundant FlcDNA sequences[75].

Alignment of FlcDNA with genomic DNA from the same region enables the determination of gene structure, including intron and exon boundaries which help to establish the reading frame and predicted amino acid sequence.

1.8.4 ***Genetic mapping populations***

1.8.4.1 *Bi-parental*

As the name suggests bi-parental mapping populations are formed from the cross-hybridisation of two usually diverse parental lines segregating for phenotypes of interest. If the trait of interest is qualitative, phenotypic data is combined with genome-wide genotypic data to calculate the relative position of loci underlying the trait of interest on a genetic linkage map, with markers which are more frequently inherited

together showing tighter linkage. In the case of quantitative traits such as yield, a QTL mapping approach is taken; briefly, a comparison between the phenotypic means of groups of markers is used to identify markers which are associated with change in mean of a particular phenotype. Numerous studies have used these techniques successfully to identify loci[69,77] and in some cases genes within barley[13,78,79] but to achieve high resolution, mapping populations require large numbers of individuals and /or high numbers of polymorphic markers to maximise the likelihood of identifying recombination events surrounding the trait of interest. Whilst this can help locate regions of interest, results are often specific to the population studied and therefore further genotyping/DNA resequencing is likely to be required to apply the results to other populations.

1.8.4.2 *Association mapping*

In contrast to bi-parental mapping, GWAS (genome wide association analysis) takes advantage of the recombination events that have occurred within a pool of germplasm (this could be a natural population or a collection of breeding lines) to identify regions of significant linkage disequilibrium (LD) between markers and traits. LD relies on closely linked markers being inherited together but requires sufficient numbers of markers for its detection. The ability to detect LD has been facilitated by the availability of genome-wide, SNP dense, genetic linkage maps.

The advantages GWAS provides over bi-allelic mapping populations is that allele segregation can be observed in a much wider pool of germplasm, enabling more general conclusions about marker trait associations to be drawn; this widens applicability.

However, there are a number of factors which can influence GWAS associations and if not taken into account can result in false associations between SNP and phenotype. Population structure is one of these factors; ideally for GWAS the population should be unstructured, i.e. all individuals unrelated, minimising the chance of alleles being associated with traits because by chance they were co-inherited, rather than being in LD. Due to the minimal domestication events and high levels of inbreeding, this is not the case for barley. Nevertheless, during the breeding of cultivars, a large number of recombination events have served to reduce the persistence of LD across chromosomes[80]. In addition, relatedness can be accounted for within GWAS through the inclusion of principal components of marker relatedness (Eigenstrat)[81] or kinship matrices as cofactors[82].

Other factors influencing the strength of associations again include marker density and population size. However, association analysis will only identify high frequency allele:trait associations; typically, SNPs with a minor allele frequency of less than 10% in the population under study are removed from the population prior to analysis to minimise spurious associations.

1.9 *The genetics underlying barley row-type*

1.9.1 *Early crossing investigations*

As mentioned previously crosses between two-rowed and six-rowed cultivated barley in the late 1800s and early 1900s identified the fact that when crossed together the six-rowed character behaved in a recessive manner to the two-rowed character, with a 3:1 segregation ratio suggesting a single locus underlying the change from two-rowed to six-rowed[52,83]

Phenotypic analysis of the F_3 generation of the cross two-rowed Svanhals × six-rowed Manchuria found that in addition to the typical two-rowed and six-rowed phenotypes further intermedium phenotypes with lateral spikelets of varying sizes, levels of awnedness and degrees of fertility were present. To account for this variation in phenotype a two-factor model was proposed[84]:

—Factor A was associated with lateral awn development and the switch from two-rowed to six-rowed. Factor A was concluded to be epistatic to factor B when homozygous six-rowed.

— Factor B was a fertility factor but its highest levels of fertility, in the absence of six-rowed factor A, were less than that of a six-rowed.

Standardisation of nomenclature later identified Factor *A* as *VV*/*vv* (two-rowed/six-rowed) and Factor *B* as *II*/*ii* (Intermedium/non-intermedium)[85]. The intermedium phenotype was also later specifically defined as: "*1,2 or 3 spikelets at a rachis node setting mature seed alternatively. The central spikelet is always fertile and better developed than the lateral spikelets*"[86].

In a subsequent study, Leonard[87] identified a third allele of the *I* locus, I^h, from the two-rowed cultivar Mortoni, which was associated with intermediums with higher levels of lateral fertility compared to the *I* allele.

Through detailed investigations of the cross Manchuria × Svanhals, Woodward[88] concluded the following genotype/phenotype relationships for the V and I loci:

VVii—two-rowed phenotype, infertile lateral spikelets with rounded lemmas.

VVII—two-rowed phenotype with rounded laterals that are more developed than *VVii*, and with occasional lateral grain development.

VVI^{h}I^h—varying degrees of lateral fertility but no awn development, lateral lemmas pointed.

Vvii—low lateral fertility intermediate, partial lateral spikelet development with pointed to short awned lemmas and occasional grain development.

VvII—similar phenotype to *Vvii* but increased lateral spikelet fertility.

VvI^{h}I^h/VvI^{h}I/ VvI^{h}i—partial to full lateral spikelet fertility. *VvI^{h}I^h* particularly resembled six-rowed phenotype.

vvii, vvII, vvI^{h}I^h—six-rowed phenotype.

Through studies of the inheritance of morphological markers, loci *VV* and *II* were found to segregate independently and assigned to separate chromosomes, *VV* to chromosome 2 (2H) and *II* to chromosome 4 (4H)[89].

1.9.2 ***Barley row-type mutants***

To gain further understanding into the genes underlying row-type, in parallel to the cross-hybridisation strategies, studies using induced mutations identified further loci with altered lateral spikelet fertility in addition to the *V* and *I* loci.

1.9.2.1 *Induced mutations*

It is estimated that spontaneous genetic mutations which result in an observable phenotype arise at a rate of 1 in 10,000 plants, however, the rate of mutation can be substantially increased through the use of chemical or radiation based mutagenesis[86].

The technique of induced mutation was used primarily for the identification of novel alleles which could provide potential improvements to the barley crop. One such example is the highly successful spring two-rowed barley cultivar Golden Promise, an induced mutant of the cultivar Maythorpe. Initially radiation based sources including gamma and x-ray were used but more latterly chemical mutagens such as sodium azide

and ethyl methanesulfonate (EMS) with the advantage that they induce single nucleotide changes to the DNA sequence compared to potentially less stable deletions associated with radiation treatments. The nature of the mutagens used means that mutational events are random, therefore only those mutations which target the spike primordia cells have the potential to be propagated into the next generation ($X_{2:}$ the second generation following mutagenesis). Phenotypic screening in the X_2 and X_3 generations is used to identify homozygous mutants[90].

During the 1900s, the Swedish Seed Association in Svalöf carried out a large proportion of the induced mutagenesis studies of barley. A range of mutagens were applied to seed of several barley cultivars, principally Bonus, Foma and Kristina. From the resultant populations mutants were categorised according to phenotype with allelism testing used to identify if independent mutations were alleles of the same locus. This led to the release of the successful European cultivar Pallas, a direct mutation of Bonus[91]. In addition, this programme generated large numbers of mutants in specific classes, e.g. the eceriferum altered wax collection[92]. The mutation collections are now archived within NordGen[93] (Nordic Genetic Resources Center) and provide valuable tools to understand the genetics underlying different barley phenotypes including row-type.

1.9.2.2 *intermedium mutants*

With respect to row-type and lateral spikelet fertility, 12 classes of independent induced or spontaneous intermedium mutant were identified in the Swedish mutation studies. All intermedium mutants are recognised by the prefix *int*, with independent loci identified by different letters, the exception being those mutants found to be allelic to the *v* locus which were designated as *hex-v (*hexastichon)[28,94-105]. In a separate induced mutant study Takahashi *et al.*[106,107] identified 5 independent six-rowed loci, these loci were recognised by the *vrs* prefix with mutants allelic to the *v* locus identified as *vrs1*. Later allelism testing confirmed 4 of the 5 *vrs* loci to be allelic to the *intermedium* loci[108]. In addition the *I* intermedium locus was found to be allelic to *Int-c*[109]. All intermedium loci are recessive, with the exception of some alleles of *int-d* and *hex-v* which have a heterozygous phenotype (pointed-laterals) and are therefore classified as semi-dominant[110-112]. The characteristic homozygous recessive phenotypes of the independent intermedium loci are described in Table 1 and images of representative spikes in Figure 9.

Figure 9: Representative spikes of the *intermedium* mutant collection.
(Photographs adapted from U.Lundqvist)

Table 1: Phenotypic descriptions of the 12 induced *intermedium* mutants
(Phenotypes as described in the Barley Genetics Newsletter Stock Descriptions[113])

Mutant allele	Phenotype
vrs3/ int-a	The upper half of the spike appears six-rowed, but lateral spikelets in the lower half are reduced in size. The basal portion of the spike appears two-rowed. Awns on the lateral spikelets range from normal near the top, to awnletted or pointed in the middle, to awnless at the base of the spike. Seeds may develop in lateral spikelets of only the upper two-thirds of the spike. Central spikelets often have double awns, one on the lemma and one on the palea. The rachilla may be deformed.
int-b	All lateral spikelets are reduced in size, and the lemma awn is short or reduced to a pointed tip. Developmental irregularities occur commonly in the lower portion of the spike Only lateral spikelets in the middle of the spike set seed Mutant plants have reduced vigour and tillering.
int-c/ vrs5	The lemma apex of lateral kernels is rounded or weakly pointed, awnless or short-awned Different alleles show varying degrees of lateral fertility
vrs1/ hex-v/ int-d	Six-rowed phenotype. The lemma awn of lateral spikelets will vary from 3/4 to nearly as long as those of central spikelets. *hex-v* and *vrs1* alleles have fully fertile lateral spikelets. *int-d* alleles vary in the level of fertility of the lateral spikelets from infertile to fully fertile but the grain produced are smaller compared to *hex-v/vrs1* mutants.
int-e /vrs4	Enlarged lateral spikelets that may set seed in the upper two-thirds of the spike The rachilla may be deformed by partial formation of an extra spikelet The awn size on lateral spikelets range from a pointed apex to 3/4 normal length Some alleles produce numerous extra spikelets at the base of the lateral spikelets and on the rachilla
int-f	The spike appears six-rowed, but the lateral spikelets are much smaller, less than half the size of the central spikelets. Lateral spikelets are pointed and often have short awns. Seed set occurs in the lateral spikelets in the upper third of the spike. The base of the spike has shortened rachis internodes.
int-h	Lateral spikelets are enlarged and have slightly pointed apex, but do not set seed. Mutants show early heading and have an elongated basal rachis internode Spikes appear lax with shortened rachis internodes at the base. Fusion of some spikelets results in double and occasionally triple kernels
int-i	Lateral spikelets are enlarged and slightly pointed at the apex, but do not set seed. The tip of the spike may have very short rachis internodes and appears very dense or fasciated

int-l/ lnt-l	Lateral spikelets are enlarged and pointed Tiller number is reduced to 2 to 4 per plant. Spike may have irregular rachis internode lengths and is relatively short Culms are thick and stiff and leaves are dark green
int-k	The spike is short and dense. Lateral spikelets are enlarged and the apex is pointed, and they occasionally have a short awn. Seed set does not occur in lateral spikelets and central spikelets are semi-sterile.
int-m	The spike is very short and has irregular rachis internode lengths. Lateral spikelets are enlarged and pointed, but they do not set seed. Spikelet density at the base of the spike is increased. Rachis internodes at the tip of the spike are very short, and the spike appears to have two or three fused or fasciated terminal spikelets. Tillering of *int-m* plants is increased
vrs2	The spike appears similar to the ordinary six-rowed spike. The lateral spikelets on the upper and lower portion of the spike are reduced in size and less fertile. Laterals are pointed with reduced awn length or awnless.

1.9.3 ***Chromosome localisation of intermedium loci***

In all cases at least one allele from each of the intermedium loci is represented in the Bowman near Isogenic line collection[114]. This collection of 881 barley near isogenic lines comprises morphological mutants which have been recurrently back-crossed, with selection for the mutant phenotype, into the North American spring two-rowed cultivar Bowman[115]. This population aids comparison between mutants as genetic differences due to different parental cultivars used in deriving the mutants are minimised.

Whilst some of the intermedium mutants had been assigned to linkage groups through linkage studies with morphological markers[89,116], genotyping of the Bowman Isoline collection with 1536 SNP markers from the Barley Oligo Pooled SNP Array1 (BOPA1)[65] provided a more defined chromosomal location for most. A series of linked polymorphic SNP between Bowman and the isoline were interpreted as representing the chromosomal segment in which the mutant locus was located (Figure 10).

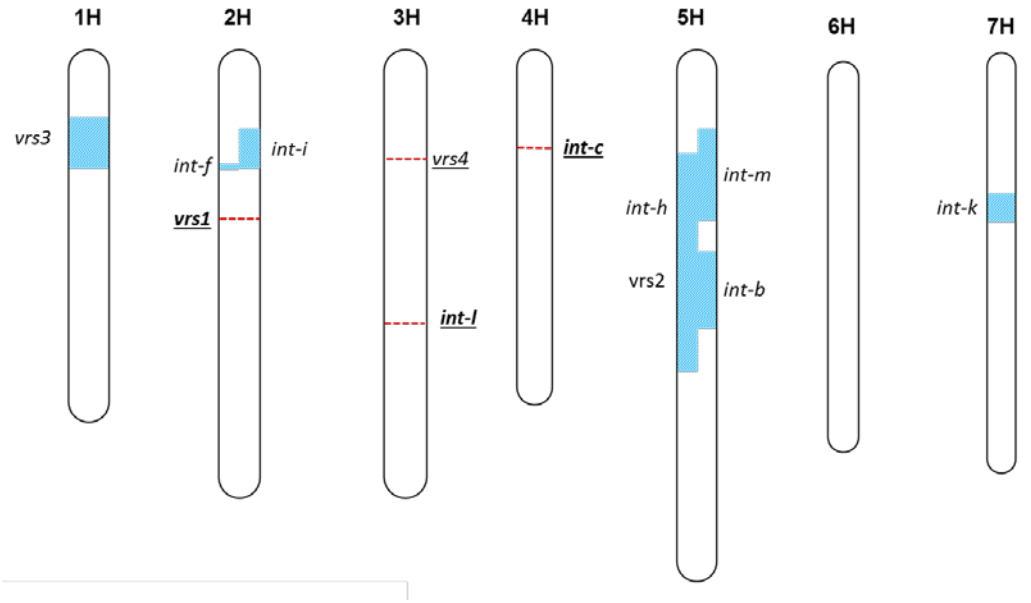


Figure 10: The chromosomal locations of the *intermedium* mutants
Loci which have been identified to the gene level are underlined and the position of the locus marked by a dashed line. In the case of loci for which the gene is still to be identified, approximate chromosome locations as identified by Druka et al[114] are represented by blue boxes.

1.9.4 ***Previously identified row-type genes***

Of the 12 *INTERMEDIUM/VRS* loci, *VRS1*, *INT-C*, *INT-L* and during the course of this study, *VRS4* have been characterised to the gene level.

1.9.4.1 *VRS1*

Through a combination of AFLP marker mapping[117] and BAC resequencing, Komatsuda *et al.*[78] identified *Vrs1* as an HD-ZIP1 transcription factor, *HvHox1*. Inactivation of *HvHox1* releases suppression of lateral spikelet development at the spikelet primordia growth stage resulting in the enhanced development of lateral spikelets and the six-rowed phenotype. The transcription factor is comprised of two conserved domains; the Homeodomain, responsible for target DNA binding and the Leucine Zipper domain which enables dimerisation between HD-ZIP1 peptides.

Whilst the genetic pathway by which de-repression of the lateral spikelets occurs is yet to be identified, in other species HD-ZIP1 transcription factors have been implicated in the control of cell differentiation and division particularly in response to drought and light stress[118].

Initial investigations identified three naturally occurring six-rowed alleles, *vrs1*.a1, *vrs1*.a2 and *vrs1*.a3 and two two-rowed alleles *Vrs1*.b2 and *Vrs1*.b3 within cultivated barley suggesting at least three separate introgressions of the six-rowed phenotype into

domesticated germplasm[78]. Further resequencing of *VRS1* within 102 accessions of the Barley CAP core collection identified a fourth six-rowed allele *vrs1*.a4[119]. The *vrs1.a*1 and *vrs1.a*2 alleles are predicted to result in truncated HvHOX1 protein sequence, the *vrs1*.a3 allele results in an amino-acid substitution within the Homeodomain and the *vrs1.a*4 allele is the result of mutation within the 5'UTR. In studies of North American and European germplasm *vrs1*.a1 and *vrs1*.a3 have been identified as the predominant alleles within cultivated six-rowed germplasm and *Vrs1*.b3 the predominant allele within cultivated two-rowed germplasm[72,119].

Vrs1 is believed to have originated from the duplication of another HD-Zip locus, *HvHox2*, 27 cM proximal on chromosome 2H[120]. Sequence alignment identified a 300bp insertion in the *Vrs1* sequence relative to *HvHox2*, resulting in the introduction of the *Vrs1* stop codon which is absent in *HvHox2*. VRS1 amino acid sequence is therefore predicted to be truncated compared to HvHOX2. An across species analysis found the non-truncated form to be the conserved form and it was therefore concluded that *Vrs1* is the product of duplication of *HvHox2* and not *HvHox2* from *Vrs1*. This duplication event is thought to be specific to barley and as such only orthologues of *HvHox2* have been identified and not *Vrs1*.

The exact role of *HvHox2* remains to be determined, but its expression profile suggests a more general role compared to *Vrs1*[121]. Both two-rowed and six-rowed cultivars share alleles making *HvHox2* unlikely to be specifically linked to row-type[120]. *OsHox14* has been identified as the rice orthologue of *HvHox2*[120,122]. Over expression of *OsHox14* in Arabidopsis resulted in a reduction in petal and sepal size and partial sterility but mutant phenotypes of *OsHox14* in rice are yet to be identified[123].

1.9.4.2 *INT-C*

Using a GWAS of the row-type of elite UK breeding lines and genotyping with pilot SNP from the Barley Oligo Pooled Arrays (pOPA) Ramsay *et al.*[72] identified a highly significant association for row-type on chromosome 4H. Taking advantage of the conserved gene order with rice chromosome 3 in the region, the *TCP* gene (named after the genes first identified in the family *teosinte branched 1, Cycloidea*, and *Proliferating Factors 1* [124]) *OsTb1*, an orthologue of the maize *teosinte branched 1* gene (*ZmTb1*) was identified as a candidate for *Int-c*; this was confirmed by sequence variation in this candidate amongst members of the *int-c* induced mutant allelic series[103] produced from the Swedish induced mutation studies.

The TCP transcription factors are characterised by the presence of a basic helix-loop-helix domain, enabling DNA binding and protein-protein interaction. They have been identified across a range of plant species and been implicated in the control of cell proliferation during plant development. However, the mechanisms by which they modulate gene transcription are still poorly understood[124].

In maize Zm*Tb1* has been identified as one of the genes underlying the domestication of maize from its wild progenitor teosinte[16]. Maize has reduced branching, with one main axis of growth in comparison to the highly branched phenotype of its progenitor teosinte, this change in architecture by suppression of axillary bud outgrowth has been found to be controlled by *ZmTb1*[125]. In addition to changes in tillering, the maize *Tb1* allele increases the number of kernels per row on the maize cob compared to teosinte[16].

A region 41kb upstream of the *Zmtb1* cDNA has been implicated in the regulation of expression of Zm*tb1*[126]. In particular, a region spanning -58kb to -69kb upstream was found to be crucial for the increased tillering and decreased kernels per row phenotypes observed in teosinte compared to maize. This study also suggested that the various control elements in the region could be directing tissue specific expression of *Zmtb1*[126].

OsTB1 in rice has been shown to negatively regulate lateral branching, over expression led to rice plants with very reduced numbers or no tillers and reduced *OsTb1* expression in the allelic *Fine-culm* (*Fc1)* mutant resulted in plants with increased tillers but finer culms[127]. Takeda *et al.*[127] postulate that *OsTB1* suppresses lateral branching subsequent to the formation of apical buds as although a reduction in tillering was observed, when dissected tiller buds were still present.

Overexpression of *tb1* in wheat was found to result in an increased number of spike rachis nodes, however, the spikelets formed were smaller with incomplete florets and sterility at the top and base of the spike. Other phenotypes observed included a reduction in tiller number and a decrease in plant height[128].

Ramsay *et al.*[72] also demonstrated that the Bowman isoline *int-c*.5 mutation led to an increase in tillering compared to Bowman in juvenile plants, suggesting that its selection in six-rowed cultivars may assist in counteracting the reduced tillering associated with *vrs1* although the effect on fertile tillers was not tested.

Within cultivated two-rowed and six-rowed germplasm , two alleles of *INT-C*, *Int-c*.a and *int-c*.b have been identified so far. The *Int-c*.a and *int-c*.b alleles differ by 3 non-

synonymous substitutions and a 6bp insertion in *Int-c*.a compared to *int-c*.b. The nucleotide sequence remains in-frame following the insertion and therefore it is not clear what effect the insertion has on the overall conformation and functionality of *INT-C.*

In the majority of cases the *Int-c*.a allele is associated with six-rowed alleles of *vrs1* and the *int-c*.b allele associated with two-rowed alleles of *Vrs1*[72] confirming the two-factor hypothesis proposed by Harlan[84] and Leonard[87]. In a six-rowed *vrs1* background, the *int-c*.b allele results in smaller lateral grain, compared to the *Int-c*.a allele. In a two-rowed *Vrs1* background the *Int-c*.a allele results in enlarged pointed lateral spikelets, with anthers but no grain-fill. However, it still remains unclear how the *Int-c*.a allele relates to the *I* and I^h alleles identified in earlier cross-hybridisation studies.

1.9.4.3 *INT-L*

int-l or as its synonym *low number of tillers-1* (*lnt-1*), suggests has a reduced tillering phenotype, with fewer primary and no secondary tillers. Spikes are also affected with lateral spikelets enlarged but infertile and entire spikelets missing in some cases. Dabbert *et al.*[129] identified *Jubel-2,* a bell-like homeodomain protein, as a likely candidate for the *lnt-l* locus. Reduction in the formation of tiller buds in the *lnt-l* mutant suggested that *Jubel-2* is involved in the formation of axillary buds rather than suppressing axillary bud outgrowth as is the case of *Tb1* locus[129].

1.9.4.4 *VRS4*

Vrs4 is characterised by a branched inflorescence phenotype, caused by the growth of supernumerary florets and spikelets [99,107,130]. Through map based cloning Koppolu *et al.*[130] identified the gene underlying this locus to be a LOB domain transcription factor, orthologous to the maize gene *Ramosa-2 (Ra2).* The role of *Vrs4* in inflorescence development is the maintenance of determinacy in the developing inflorescence meristem with a similar role reported for *Ra2* in maize[131].

Resequencing in cultivated germplasm identified 8 alleles of *HvRa2,* one major and seven minor, however unlike *Vrs1* and *Int-c* no single allele was associated with a particular row-type. Expression analysis found *Vrs1* highly down-regulated in a *vrs4* mutant background; implicating *Vrs4* in a role mediating the expression of *Vrs1*[130]. It remains unclear whether this mediation is through direct interaction with *Vrs1* or via an intermediary.

1.10 *Research objectives*

Whilst four genes influencing the lateral spikelet development in barley have been identified, a further eight remain to be identified. It is not clear if or how these loci interact and the genetic pathway controlling barley inflorescence architecture remains to be determined overall, a crucial step in understanding the potential for improvements in six-rowed barley yield potential and grain size distribution.

Of the remaining unidentified intermedium loci, *vrs3* has the largest number of independent induced mutations, 32 in total. In studies of the interactions between intermedium loci, Lundqvist *et al*.[132-134] suggested that *int-a* alleles combine well with *hex-v, int-d* and *int-c* mutants to produce regular six-rowed spikes with enhanced lateral grain size. However, due to the lack of available DNA markers the identification of genotype combinations relied on phenotypic characterisation to infer the underlying genotype which was not always easy to determine particularly when segregation occurred at three and four genetic loci [133,134].

The aims of this study were to further the current understanding of the genetics underpinning barley row-type and investigate the potential for improvements to the current cultivated six-rowed germplasm by:

—Further refinement of the position of the *VRS3* locus on chromosome 1H and thereby identification of potential candidate genes.

—Use of the *int-a* induced mutant allelic series to confirm the identity of the gene underlying *VRS3* and investigate differences in phenotype associated with the identified mutations.

—Investigation of the sequence diversity of *Vrs3* within cultivated germplasm and its conservation across plant species.

—Investigation of the potential of *vrs3* as alternative six-rowed allele within cultivated germplasm, particularly with respect to improving the uniformity in grain size between the lateral and central spikelets; either alone or in combination with allelic variants at *VRS1* and *INT-C*

—Investigation of the expression of *Vrs*3 in conjunction with the other identified row-type genes to try and establish a network of expression.

2. Mapping *VRS3*

2.1 *Introduction*

As described in Chapter 1 row–type refers to the relative fertility of the spikelets in the barley spike and is one of the major phenotypic sub-groupings of barley. In six-rowed barley all three spikelets at a rachis node are fertile and grain producing. In two-rowed barley, only the central spikelet is fertile, thereby only producing a single grain. The genetic control of the change between two-rowed and six-rowed barley has identified 12 loci associated with lateral spikelet fertility. In particular, the loci *VRS1*, *INT-C* and *VRS4* have been characterised at the gene level (Section 1.9.4). *VRS3*, the focus of this study, is a third locus which affects lateral floret fertility.

2.1.1 ***Identification of the vrs3 phenotype***

A mutant *vrs3* phenotype was first identified by two independent research groups. In Japan, the mutant Kmut213 was created through the gamma ray induced mutation of two-rowed cultivar Hakata 2. The phenotype of Kmut213 was described as:

"underdeveloped lateral spikelets only in the lower 1/4 of a head, and often by deformed rachilla"[135]

Using allelism testing with another induced mutant, Kmut27, previously identified as v2 (*vrs2*) and a six-rowed cultivar Natsudaikon-Mugi (*V, vrs1*), it was concluded that a separate locus was responsible for the change in row-type observed in the Kmut213 mutant, which led to the locus being classified as *v3* (*vrs3*)[135]. *vrs3* was further localised to barley chromosome 5 (chromosome 1H) using trisomic analysis [106], which clearly showed the absence of the mutant phenotype amongst the trisomic F2 individuals. The Kmut213 mutant has since been given the allele nomenclature *vrs3*.f[136].

As discussed in section 1.9.2.2, the 'intermedium'mutants were identified as part of a mutant breeding programme at the Svalöf Institute, Sweden[86]. They used genetic complementation testing and comparison of phenotype to classify alleles into different locus groups.

The Svalöf Institute identified 31 *intermedium-a* (*int-a*) mutants, publishing the following generic description to describe the group:

"The upper half of the spike appears six-rowed, but lateral spikelets in the lower half are reduced in size. The basal portion of the spike appears two-rowed. Awns on the lateral spikelets range from normal near the top, to awnletted or pointed in the middle, to awnless at the base of the spike. Seeds may develop in lateral spikelets of only the upper two-thirds of the spike. Lateral spikelets may be pedicellated in some stocks. Central spikelets often have double awns, one on the lemma and one on the palea. The rachilla may be deformed"[136].

Exact chromosomal location of a Mendelian character was problematic at the time the intermedium mutants were first classified as DNA markers had yet to be identified. Geneticists had to rely on classical linkage tests to other morphological or biochemical marker stocks and compare the observed and expected frequencies of parental and recombinant classes to place new marker loci within an existing framework. For barley, sets of cytologically characterised translocation stocks were also available and the reduced fertility of such stocks could also be used as a marker to check for significant linkage with a translocation break-point. Therefore the study of the inheritance patterns of *vrs3* with other morphological and translocation stocks markers in a series of tester crosses was used to locate the locus near the centromere of the short arm of chromosome 5 (1H)[89].

Subsequent genetic complementation testing between the *int-a* allelic series and *vrs3* found no rescue of the wild-type phenotype amongst F_1 plants from the crosses, concluding the two classes of mutants to be allelic[108]. Throughout this thesis the locus will be referred to as *VRS3* unless referring to specific mutant alleles.

2.1.2 *vrs3 mutants in Bowman near isogenic line collection*

With respect to the *vrs3* locus, the Bowman Isolines contain two introgressed alleles, the previously described *vrs3*.f (BW902) from Hakata-2 and *int-a*.1 (BW419). *int-a*.1 is an X-ray induced mutant of the two-rowed spring cultivar Bonus[136]. The pedigrees of Hakata-2, and Bonus are very different; Hakata-2, a Japanese cultivar from the cross Golden-Melon/Prior, was developed by the Nihon Brewery in 1934[137]. Bonus, a

Swedish cultivar, from the cross Maja//Seger/Opal was released by the Svalöf institute in 1950[138].

2.1.3 ***Previous genetic mapping of VRS3***

Previously genotyping of the *vrs3* Bowman isolines with the BOPA1 set of SNP markers, confirmed chromosome 1H as the location of the *vrs3* introgressed mutant segment. In the case of BW902 the introgressed segment was represented by 16 polymorphic SNP on chromosome 1H covering a distance of 20.25 cM and for BW419 12 SNP spanning a distance of 7.7 cM (IBGSC Map)[114], suggesting *VRS3* to lie within a region of chromosome 1H predicted to contain approximately 860 genes[139].

Although these intervals give a good indication of the possible position of the mutant locus, they are delimited by polymorphic SNP between Bowman and the Bowman isoline. The possibility therefore arises that the mutant introgressed interval could extend further but potential monomorphism at SNP loci between Bowman and the mutant genetic background means that this would be undetected.

2.1.4 ***Chapter summary***

Through the detailed phenotyping of four *VRS3* bi-parental mapping populations, comprised of two different *vrs3* mutants (BW419 and BW902) crossed with two two-rowed spring barley cultivars (Bowman and Barke) in polytunnel and field environments, I aimed to test the hypotheses that:

—the *vrs3* phenotype observed is affected by the genetic background.

—the extent of the *vrs3* phenotype differs between different *vrs3* mutations.

—using multiple different bi-parental mapping populations maximises the chances of identifying informative recombination events in the mapping of *VRS3*.

—the gene underlying *vrs3* lies in an interval on the short arm of chromosome 1H.

2.2 *Materials and methods*

2.2.1 *Germplasm resources*

Druka *et al*.[114] created the four F_2 populations used for this mapping study by crossing the Bowman barley isolines BW419 (*int-a*.1/6*Bowman) and BW902 (*vrs3*.f/6*Bowman) to two barley cultivars Bowman and Barke.

The parent cultivars used in this mapping study were:

Bowman a North-American two-rowed spring cultivar, bred by North Dakota State University from the cross (Klages//Fergus/Nordic/3/ND1156/4/Hector) and first registered in 1985.[115]

Barke a German two-rowed spring malting barley, bred by Saatzucht Josef Breun GdbR, from the cross Libelle/Alexis, first registered in Germany in 1995.

Crosses were carried out using the "egg-top" method of hybridisation[140]. Briefly, at Zadoks growth stage 45 (boot swelling and anthers green) [22] spikes are emasculated by cutting off 1/3 of the spikelet top and removal of the three exposed anthers using forceps. Emasculated spikes are covered with glassine bags to prevent unintended cross-pollination and left for two to three days to allow the floret to become receptive to pollen.

The pollen donor spikes are similarly treated but at a slightly later stage when the anthers are yellow and about to dehisce. When the florets are cut open, the anthers extrude and dehisce; the emasculated spike can then be pollinated by tapping of the dehisced spike over the receptive florets. Pollinated spikes are re-covered with the glassine bag, with the beginnings of seed set visible in 3-4 days.

The resultant F_1 seeds were grown to full maturity in a heated and lit glasshouse. Family structure was retained, with all spikes from a single F_1 plant harvested together to enable the identification and elimination of any selfed female seed from the original cross.

2.2.1.1 *Population and individual line nomenclature*

The nomenclature used to identify the mapping populations is shown in Table 2:

Table 2: The nomenclature used to identify individual F_2 mapping populations

Population Identifier	Cross
BW419Bo	Bowman/BW419
BW419Ba	Barke/BW419
BW902Bo	Bowman/BW902
BW902Ba	Barke/BW902

Individual F_2 plants were segregated into families, i.e. all F_2 plants grown from seed originating from the same F_1 plant were considered a family.

Individual F2 plants within a population were identified using the following format:

Population Identifier-Family Number-Individual Number

e.g. BW419Bo-1-01 is F2 plant number 01 from F_1 plant number 1 of the BW419/Bowman cross.

2.2.2 ***F_2 population polytunnel experiment***

2.2.2.1 *Experimental design and planting*

440 F_2 individuals, representing 11 families, each of 10 individuals, from the following four populations BW419Bo, BW419Ba, BW902Bo and BW902Ba were grown in pots in a poly-tunnel during the spring/summer of 2010 at Mylnefield Farm, Dundee (Figure 11). Single seed were sown in late March (26/03/10) in a 23 cm diameter pot containing a compost/Intercept granular insecticide mix suitable for growing cereals. Pots were irrigated, initially by overhead watering to thoroughly wet the soil, and subsequently from capillary matting kept moist by a drip system. Rodents ate 189 plants, across all four crosses and these had to be re-sown; residual seed from each population were pre-germinated into multicellular trays within the glasshouse (22/04/10). Seedlings were transplanted into the polytunnel once they had reached 1 true leaf, GS11 and judged to be less susceptible to rodent attack. A prophylactic fungicide programme of Opus Team (1 ml/l) at 66 days after sowing and Fortress (1 ml/l) at 78 days after sowing was applied to the whole polytunnel to prevent the development of common foliar pathogens

of barley. Prior to stem extension, plastic nets in the form of a sleeve were put around the pots to support the plants and avoid any lodging whilst allowing light to reach the plant (Figure 11).

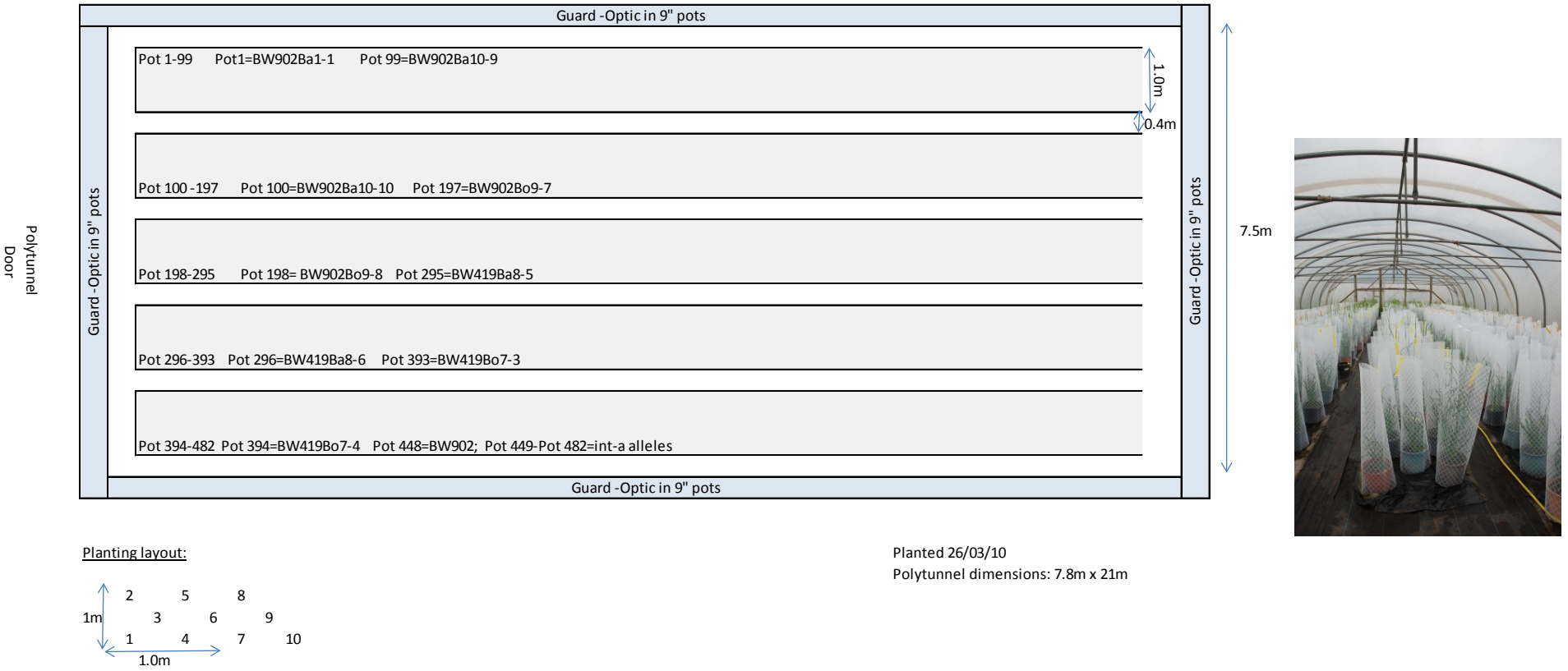


Figure 11: Detailed field plan of F_2 populations within polytunnel
(inset photograph shows actual polytunnel layout)

2.2.2.2 *Tissue sampling*

50—100 µg of young leaf tissue was harvested from each plant in the polytunnel at approximately GS25. Tissue was snap frozen in liquid nitrogen and stored at -80^{o}C until DNA extraction.

2.2.2.3 *Phenotyping*

Plants were scored for spike emergence, specifically the date when 50% of tillers per pot had spikes which had emerged half-way above the flag-leaf auricles (GS55). This was done by surveying the whole experiment three times weekly until the last plant had completed spike emergence. At the early dough stage of grain ripening (GS83), one representative spike was removed from each plant, photographed as an entire spike, representative spikelets from the top (T), middle (M) and bottom (B1 and B2) of the spike removed and also photographed. Row number, awn length, grain-fill, awned palea, abnormal and multi-florous characteristics were recorded on central and lateral florets. Spikes were classified as *vrs3* mutant when evidence of lateral spikelet grain-fill and lateral spikelet awn development were found. Spikes were considered wild-type when all lateral spikelets were infertile and had lemmas which were reduced in size with rounded tips and no awn development, characteristic of two-rowed barley. Spikelets were defined as awnletted when the lemma awn was less than 2.5cm. Further phenotyping with respect to height and tiller number was carried out pre-harvest. Three measurements of height were recorded as distance from the soil surface to the collar, spike tip and average awn tip. Tiller number was recorded as the total number of fertile stems per plant.

2.2.3 ***F_3 field trial***

2.2.3.1 *Field trial design*

Approximately 20 F_3 seeds derived from each F_2 plant that had been genotyped with the BeadXpress SNP set, i.e. approximately 100 F_2 progenies from each of the six populations, were sown in paired-rowed 0.5 m^2 plots. The plots were sown on 05/04/11 with a Wintersteiger Seedmatic drill and arranged in a row and column replicated design that was 40 plots deep and 30 plots wide. The parents and selected *vrs3* mutants were included to make a total of 600 entries.

2.2.3.2 *Agronomy*

The plots were combine drilled with a compound fertiliser at a rate of 77 Kg N/ha 14 Kg P/ha and 49Kg K/ha. A further 33 Kg N/ha was applied as a top-dressing when the

crop reached GS22. At GS22 a fungicide mix of Proline (0.8 l/ha) + Comet 200 (0.75 l/ha) + Bravo 500 (1.0 l/ha) + Talius (0.15 l/ha) was applied to prevent infection by common foliar pathogens. At GS39 (T2), a further fungicide regime of Proline (0.8 l/ha) + Comet 200 (0.75 l/ha) + Bravo 500 (1.0 l/ha) was applied. When the majority of the plots were ripe (GS91), each plot was harvested as a bundle of whole plants from the beginning to-mid September, dried, and stored until post-harvest phenotyping was carried out (1.2.4.4).

2.2.3.3 *Field phenotyping*

At flag leaf emergence (GS39), one plant per plot was selected at random, tagged and scored for tiller number. Spike emergence date was scored tri-weekly on a whole-plot basis as the date when 50% of the tillers within a plot had spikes that had emerged 50% from the boot (GS55). At the late-milk stage of grain ripening (GS77), plots were scored for spike row-type (two-rowed, six-rowed, heterozygous (mixed)).

2.2.3.4 *Post-harvest phenotyping*

The plant tagged at GS39 in each plot (see above) was re-scored for tiller number to obtain an estimate of fertile tillers, i.e. those tillers which had fully developed spikes which had produced grain and were fully senesced on harvest (not green). Plots from crosses BW419Bo, BW419Ba, BW902Bo and BW902Ba were all re-assessed for row-type, to ensure cross-reference with field scores.

2.2.4 ***Statistical analysis of phenotypic data***

All statistical analysis was carried out using GENSTAT version 14[141].

The Pearson Chi-squared method was used to test the goodness of fit between observed and expected segregation ratios for the *vrs3* phenotype in F_2 populations. Pearson Chi-squared tests of association, in combination with a random-permutation test were used to test association between *VRS3* genotype and observed phenotypes in the BW419Ba, BW419Bo, BW902Ba and BW902Bo F_2 populations. In general, phenotypic means were generated using REML (Restricted Maximum Likelihood) with entries as fixed and blocks as random effects. In addition, incomplete blocks were fitted as random effects for the field trial if there was evidence of a significant spatial trend across the experiment. Additional terms were added for each experiment as follows:

1. **F_2 polytunnel experiment**: The effect of the re-sowing due to mice damage, parent, mutant-allele and F_2 row-type were modelled as fixed effects

3. **F_3 Bowman and Barke populations**: for the analysis of spike emergence a linear model was fitted with bloc, parent, mutant allele and row-type of plot fitted as fixed effects. The analysis of tillering used a slightly different model with respect to row-type, fitting the row-type of the specific plant in which tillering was counted.

4. Residuals were plotted against row and column coordinates to identify any spatial trends. The model was also tested with the inclusion of row and column as random factors to observe if their inclusion gave significant improvement in the model as tested by a significant reduction in the overall deviance. A probability level of <0.05 was taken to be evidence of a significant effect.

2.2.5 *Genomic DNA extraction*

Prior to DNA extraction, approximately 100 µg frozen young leaf material (sections 1.2.2.2, 1.2.3.2) was ground to fine powder using a micro-pestle. DNA extractions were then carried out using the QIAGEN DNeasy plant minikit as per the manufacturer's protocol (July 2006 version) (http: //www.qiagen.com/gb/products/catalog/sample-technologies/dna-sample-technologies/genomic-dna/dneasy-plant-mini-kit/#resources). Genomic DNA was eluted into a total volume of 50 µl with the final elution in two stages. DNA eluent from the initial stage was passed back through the column for a second time to increase final DNA concentration. DNA concentration was determined using Picogreen reagent (Invitrogen). Briefly, three replicates of the following dilution series of lambda DNA (75, 50, 25, 12.5, 6.25, 3.125, 1.5625 and 0 ng/µl) were used for standard curve determination. Picogreen reagent was diluted 1/200 with TE. 197µl of diluted Picogreen was added to 3µl volumes of DNA (both dilution series and extracted DNA of unknown concentration) in 96-well plates. After incubation at room-temperature for 5 minutes, fluorescence was determined using a fluorescence microplate reader. ASCENT software version 2.6 was used to fit the standard curve and calculate DNA concentration from fluorescence readings.

2.2.6 *BeadXpress genotyping*

The barley BeadXpress Oligo Pooled Array (XOPA1) is a genome-wide subset of 384 SNP from the original 1536 SNP of BOPA1 designed to provide a similar genome-wide assessment of different barley genotypes as BOPA1. Briefly, BOPA1 genotyping information from 171 cultivars was used to select the SNP for the XOPA1 subset. SNP included in XOPA1 were selected according to the following criteria: <5% missing

data, minimum allele frequency of ≥0.2, and even distribution across all seven barley chromosomes [142].

BeadXpress genotyping was carried out as recommended in the Goldengate Genotyping Assay for VeraCode Manual (Illumina VC-901-1001). Genotypes were scored automatically using Genomestudio software (http: //www.illumina.com/software/genomestudio_software.ilmn). Subsequent to automated genotyping, profiles for each SNP marker were scanned manually using the Genomestudio software and those which performed poorly (fluorescence values of less than 0.2) or inconsistently were removed from the analysis.

2.2.7 ***KASP genotyping***

2.2.7.1 *Selection of informative SNP for assay design*

Further SNP for allele specific assay designs were selected from the full set of 3072 BOPA1 and BOPA2 SNPs using the following criteria:

— Prior BOPA1 and BOPA2 genotyping of Bowman, Barke and Bonus had found the SNP to be polymorphic between Bowman and Bonus and/or Bonus and Barke (JHI unpublished data).

—The Muñoz-Amatriaín[143] consensus map positioned it on chromosome 1H proximal to. 11_10438 at 42.42 cM

2.2.7.2 *KASP assay design*

DNA nucleotide sequence spanning at least 50bp either side of the SNP of interest were submitted to LGC Genomics for design of allele specific PCR (KASP) reactions using their KASP-by-design service.

All genotyping reactions were carried out in 8 µl volumes comprising approximately 50 ng genomic DNA, KASP V4.0 2X Master mix with standard ROX (KBS-1016-002) and 0.01 µl KASP by design assay. Reactions were prepared in MicroAmp® Fast Optical 96-Well Reaction Plate (4366932) with MicroAmp® Optical Adhesive Film (4311971). Genotyping reactions were run on StepOnePlus™ (Applied Biosystems) qPCR machine. Cycling conditions were as follows: 20^{o}C—2minutes; 94^{o}C—15minutes; 10 touch-down cycles (94^{o}C—20sec, 65^{o}C—1 minute (-0.8^{o}C per cycle)); 32 cycles (94^{o}C—20sec, 57^{o}C—1 minute); 20^{o}C—2minutes.

2.2.8 ***Genetic mapping of Vrs3***

2.2.8.1 *JoinMap*

JoinMap 4 (Kyazma) mapping software[144] was used to produce individual population maps using all BeadXpress markers that were polymorphic within a population and including the *VRS3* row-type phenotype as a co-dominant marker. Nearly all of the 384 Beadxpress SNP have a known barley map location therefore linkage groups were determined by the LOD score at which the highest number of SNP from a chromosome were grouped together with no inclusion of SNPs from another chromosome. Subsequently, linkage groups which were known to contain markers from the same chromosome were merged to form a single linkage group for marker ordering within a chromosome.

The regression mapping algorithm of JoinMap 4, with the Haldane mapping function[145] was used for marker ordering. Ordering was carried out with the default values for all parameters apart from the maximum recombination frequency and minimum linkage LOD for inclusion, which were set to 0.49 and 0.01 respectively. JoinMap build maps in three successive rounds. In the first, any marker whose inclusion causes a jump in the overall chi-squared for the map that exceeds the threshold (default value = 5.0) is removed. Once all markers have been tried to fit in the map, a second round attempts to fit in all the excluded markers but with the same restriction as in the first round. Finally, a third round of mapping forces all remaining markers into their best position within the map but with no restrictions on the goodness of fit of that map. The exact ordering of markers produced from the third round of mapping is therefore unreliable and can lead to many double recombinants that are statistically highly improbable. In cases where 3 rounds were necessary to include all the markers in a map of the linkage group, the third round map was ignored and the results from the second round used.

The fine mapping of the region surrounding *VRS3* based upon the recombinants between its flanking markers was also done with JoinMap 4. In this case, the results from the KASP genotyping of the recombinants between the flanking markers were combined with predicted genotypes for the non-recombinants based upon their genotypes at the flanking markers to produce a map order within the region.

2.2.8.2 *Identification of regions homologous to Vrs3 in rice*

Initially, the comparative genome visualisation software Strudel (Version 1.11.02.11)[146]was used to identify the broad region of rice chromosome homologous to the interval identified on barley chromosome 1H. Subsequently, the Barley Genome-Zipper[73]was interrogated to further delimit the region of candidate rice loci. BLAST of SNP manifest sequences (http: //bioinf.hutton.ac.uk/iselect/app/) against the MSU rice genome annotation resource[147] improved homology resolution and enabled determination of the flanking candidate loci intervals.

2.2.8.3 *Candidate gene list construction*

The results from the fine mapping were used to refine the interval containing *VRS3* to an interval contained by the two new flanking markers established after adjustment to order the regions according to the rice gene order. Assuming that the gene order in rice has been conserved, all rice loci within the interval identified with putative function or expression assignations were considered as candidates for *VRS3*.

2.3 *Results*

2.3.1 F_2 *Barke and Bowman populations*

2.3.1.1 *Chi-squared analysis*

Plants in the four F_2 populations, BW419Ba, BW419Bo, BW902Ba and BW902Bo segregated for row-type, either two-rowed, with all laterals reduced and infertile or mutant *vrs3*, with lateral spikelets showing awn development and grain fill. Figure 12 shows characteristic wild-type and mutant *vrs3* spikes from each of the four populations. In populations BW419Bo and BW902Ba, 2 of the 11 F_1 families were all two-rowed; this would suggest that they are products of a self-hybridisation event in the initial cross and were therefore discounted from further analysis.

Figure 12: Representative spikes from the four F_2 populations
a) mutant BW419Ba, b)two-rowed BW419Ba, c) mutant BW419Bo, d) two-rowed BW419Bo, e) mutant BW902Ba, f) two-rowed BW902Ba, g) mutant BW902Bo, h) two-rowed BW902Bo.

A chi-square test was used to test the segregation of two and six-rowed alleles at *VRS3*. This showed no significant deviation from the expected segregation ratio of 3:1 for a dominant gene in an F_2 population confirming previous reports that a single recessive gene at the *VRS3* locus controls the phenotype (Table 3).

Table 3: The *Vrs3* : *vrs3* phenotype segregation ratios within each of the F_2 populations

Population	Total	Two-rowed	mutant	Chi-Squared (1d.f.)
BW419Ba	108	77	31	0.77 (p=0.381)
BW419Bo	88	67	21	0.06 (p=0.805)
BW902Ba	90	62	28	1.71 (p=0.191)
BW902Bo	107	76	31	0.87 (p=0.351)

2.3.1.2 *vrs3 associated phenotypes*

To further establish the expressivity of *vrs3*, central and lateral spikelets from four areas of the spike were characterised. The areas of the spike were identified as follows:

B1 (base 1)—a spikelet triplet from the lowest rachis internode.

B2 (base2)—a spikelet triplet from the second lowest rachis internode.

M (middle)—a spikelet triplet from an internode approximately half-way up the spike.

T (top)—the uppermost spikelet triplet in the spike.

Spikelet awn development and presence/absence of grain-fill was scored independently on both central and lateral spikelets at each position on the spike. Spikelet awn development was scored on a scale of 0—5 as detailed in Table 4.

Table 4: Scoring scheme for spikelet awn development

Score	Description
0	Vestigial rounded lemma
1	Enlarged pointed lemma
2	Awn less than 1 cm (awnletted)
3	Awn 1—2.5 cm (awnletted)
4	Awn 2.5—5 cm
5	Awn greater than 5 cm

In addition to variations in awn development and grain-fill, additional inflorescence related phenotypes were observed (Figure 13); these were scored on a presence/absence basis at each point of the spike.

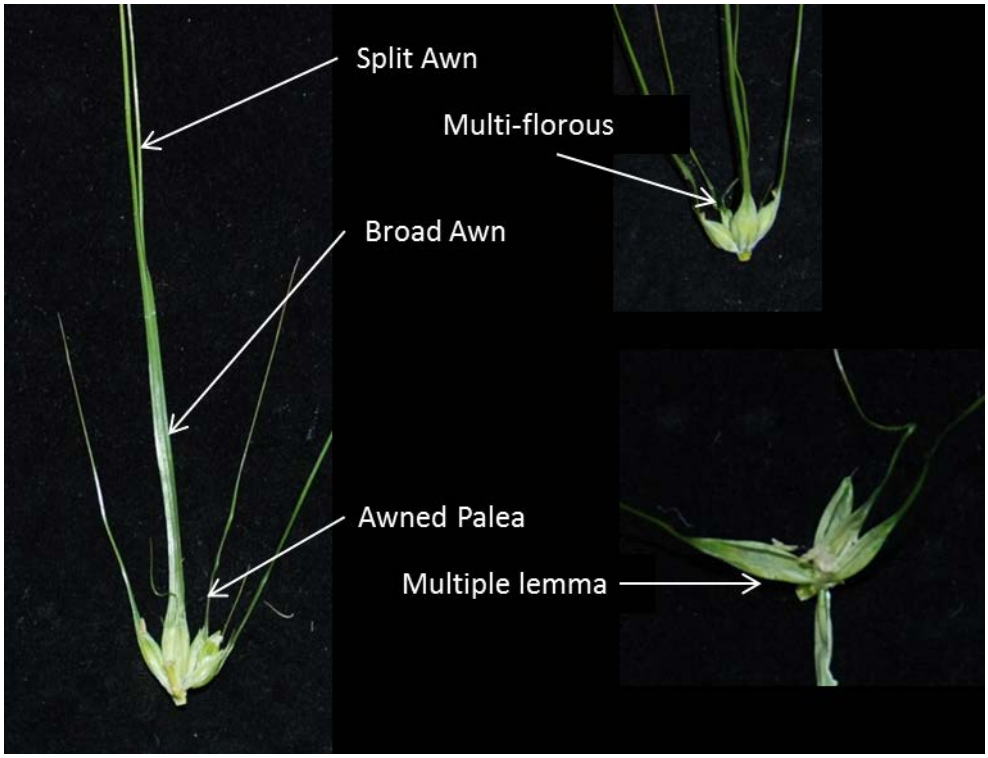


Figure 13: Examples of the *vrs3* inflorescence phenotypes

Multi-florous

The multi-florous phenotype showed supernumerary spikelets and florets. In general these arose either from the ventral furrow of the spikelet or from the node on the lemma (abaxial) side of the spikelet.

Awned Palea

Spikelets with the awned palea phenotype showed awn-like protrusions from the palea, these awns were in general shorter and thinner than the usual lemma awn.

Broad Awn

The broad awn was approximately twice as wide as the typical lemma awn.

Split Awn

The split awn was evident from approximately 2/3 the length of the awn.

Multiple Lemma

Spikelets displaying the multiple lemma phenotype had several lemmas nested within one another, in some cases up to five.

2.3.1.3 *Lateral spikelet awn development*

Across all populations and all positions of the spike there was a highly significant Chi-squared test of association between lateral awn development and row-type ($p<0.001$) (Appendix 2). The lateral spikelets of two-rowed individuals showed only rounded lemma development but more varied lateral awn development was observed in the *vrs3* homozygous individuals. Within the *vrs3* homozygous individuals, generally the greatest levels of lateral awn development were observed at the middle of the spike and the lowest levels at B1 (Figure 14).

Comparison of lateral awn development of *vrs3* homozygous individuals between those from BW419 and BW902 populations found a significant association between awn development and Bowman isoline at the B1 ($\chi^2= 16.87$, 5 d.f. $p=0.005$) and B2 ($\chi^2= 12.47$,5 d.f. $p=0.025$) internode positions. The *vrs3* homozygous individuals from the BW902 populations tended to show greater levels of awn development at these positions compared to the *vrs3* homozygous individuals from the BW419 population, indicative of a stronger mutant phenotype in BW902 compared to BW419.

A similar comparison, pooling *vrs3* homozygous individuals based on their Bowman and Barke populations, found significant associations between population and awn development at all four spike positions (Appendix 3), with those spikes from Bowman populations associated with greater awn development than those from the Barke populations. This would suggest background genetic effects influencing the awn development phenotype, with the Bowman background giving a stronger mutant phenotype.

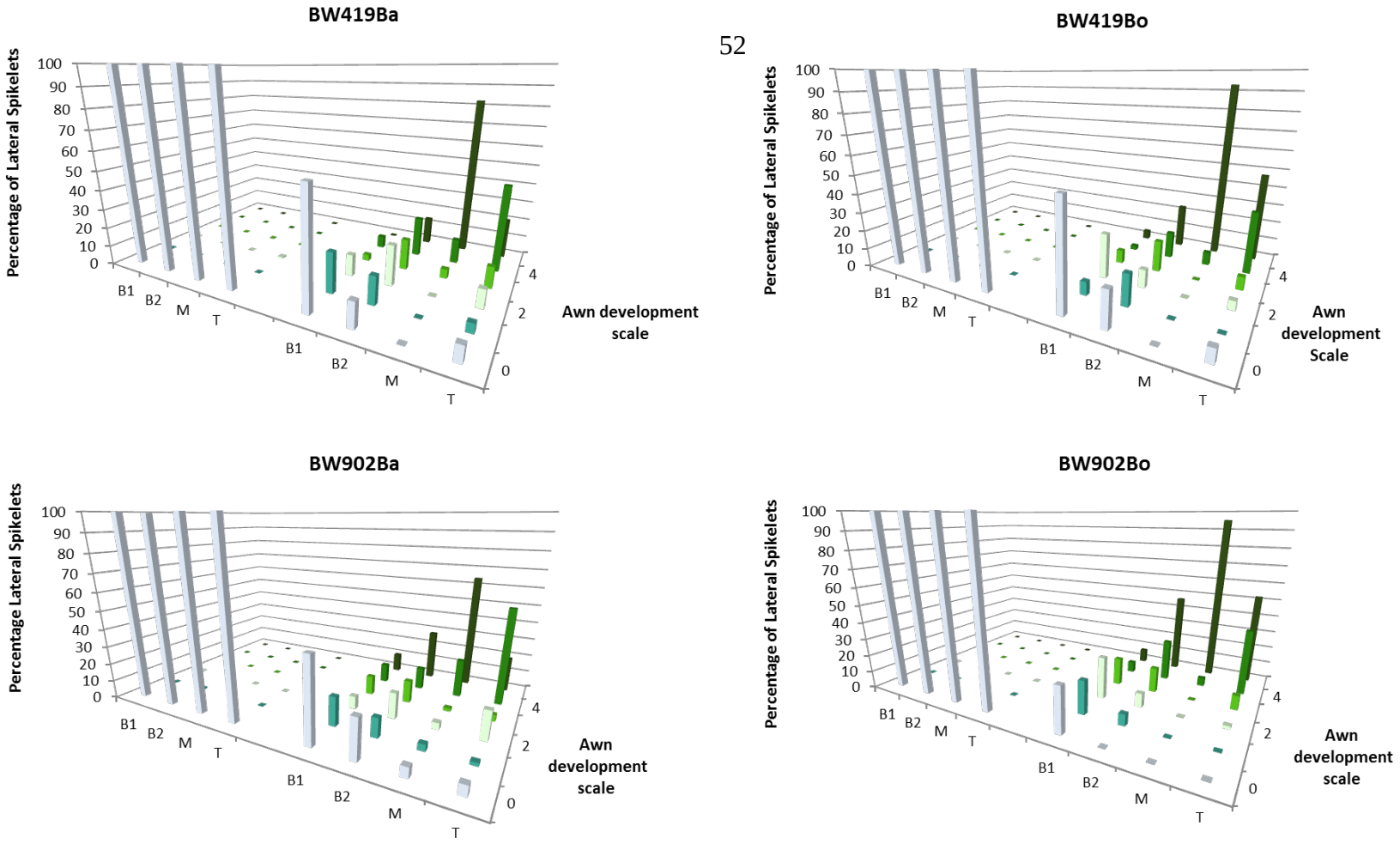


Figure 14: Comparison of lateral spikelet awn development in the four F_2 populations.

Individual graphs show the lateral spikelet awn development within a population at four regions of the spike, in both two-rowed and vrs3 homozygous spikes B1= base spikelet 1, B2=base spikelet 2 , M =middle spikelet, T=top spikelet, 0=rounded lemma, 1=pointed lemma, 2 =awn <1cm, 3=awn 1-2.5cm, 4=2.5-5cm 5=awn>5cm.

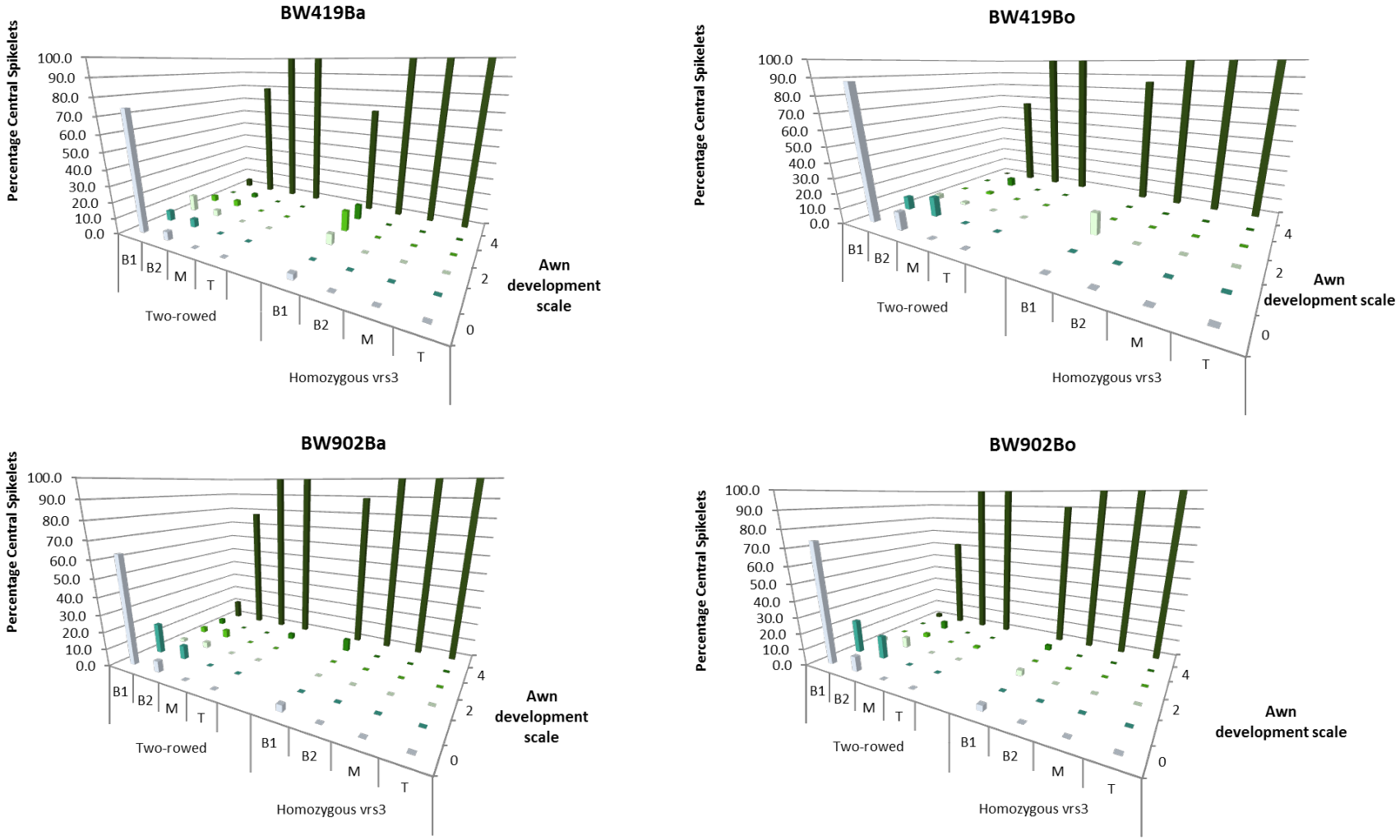


Figure 15: Comparison of central spikelet awn development in the four F_2 populations.

Individual graphs show the central spikelet awn development within a population at four regions of the spike, in both two-rowed and vrs3 homozygous spikes. B1= base spikelet 1, B2=base spikelet 2, M =middle spikelet, T=top spikelet, 0=rounded lemma, 1=pointed lemma, 2 =awn <1cm, 3=awn 1-2.5cm, 4=2.5-5cm 5=awn>5cm.

2.3.1.4 *Central spikelet awn development*

Central spikelet awn development was also found to be variable with respect to differences associated with two-rowed and homozygous *vrs3* individuals but only at the B1 and B2 spike positions. The association was most pronounced at the B1 position with all associations being highly significant ($p<0.001$; (Appendix 4)). At this position of the spike the homozygous *vrs3* individuals showed a predominance for central spikelets with awns >5 cm whereas the central spikelets from two-rowed spikes showed more variable development with a tendency for rounded lemmas (Figure 15) suggesting *Vrs3* influences central awn development at these positions.

At the B2 position, the difference between spikes of the different row-types was less but nevertheless still significant (Appendix 4). At this position the homozygous *vrs3* central spikelets all had awns >5 cm; the awns of the two-rowed spikes showed a pre-dominance of this category too but also more variable awn development with individuals represented in the majority of the five categories.

Analysis of central spikelet awn development at the B1 spike position of homozygous *vrs3* individuals found no significant association with Bowman Isoline or parent cultivar. However, the same pooling strategies in the two-rowed individuals showed significant differences in the central spikelet awn development; those two-rowed spikes originating from Barke crosses showed greater awn development than those from Bowman at the B1 position (Appendix 5) suggesting that the variation in central spikelet awn development observed cannot be fully attributed to the *VRS3* locus.

2.3.1.5 *Spikelet grain-fill*

By definition across all four populations, the two-rowed individuals showed no lateral spikelet grain-fill. Lateral spikelet grain-fill was observed in *vrs3* homozygous spikes, but the levels varied depending on population and the position of the spikelet on the spike (.Figure 16).

At the B1 position, no significant association was found between row-type and lateral spikelet grain-fill in the BW419Ba (χ^2=2.5 ,1d.f., p=0.293) and BW419Bo (χ^2=6.45, 1d.f., p=0.057) populations suggesting that at this position the BW419 homozygous *vrs3* spikes resemble a two-rowed phenotype. This was not the case for BW902, where significant associations between lateral spikelet grain-fill and row-type were identified at B1, BW902Ba (χ^2=11.39,1d.f., p=0.002) and BW902Bo (χ^2=9.99, 1d.f., p=0.007) suggesting the *vrs3* phenotype persists further down the spike in the homozygous *vrs3* spikes from BW902 compared to those from BW419.This is consistent with the findings in difference in lateral spikelet awn development between homozygous *vrs3* individuals from BW902 and BW419 (section 2.3.1.3).

Central spikelet grainfill also appeared to be influenced by *VRS3*, particularly with respect to grain-fill at the B1 position; homozygous *vrs3* spikes showed a higher level of central grain-fill than two-rowed classes ($p<0.001$) (Appendix 6). The association with row-type is greatest in the crosses to Bowman suggesting background genetic effects are also influencing the extent of the phenotype.

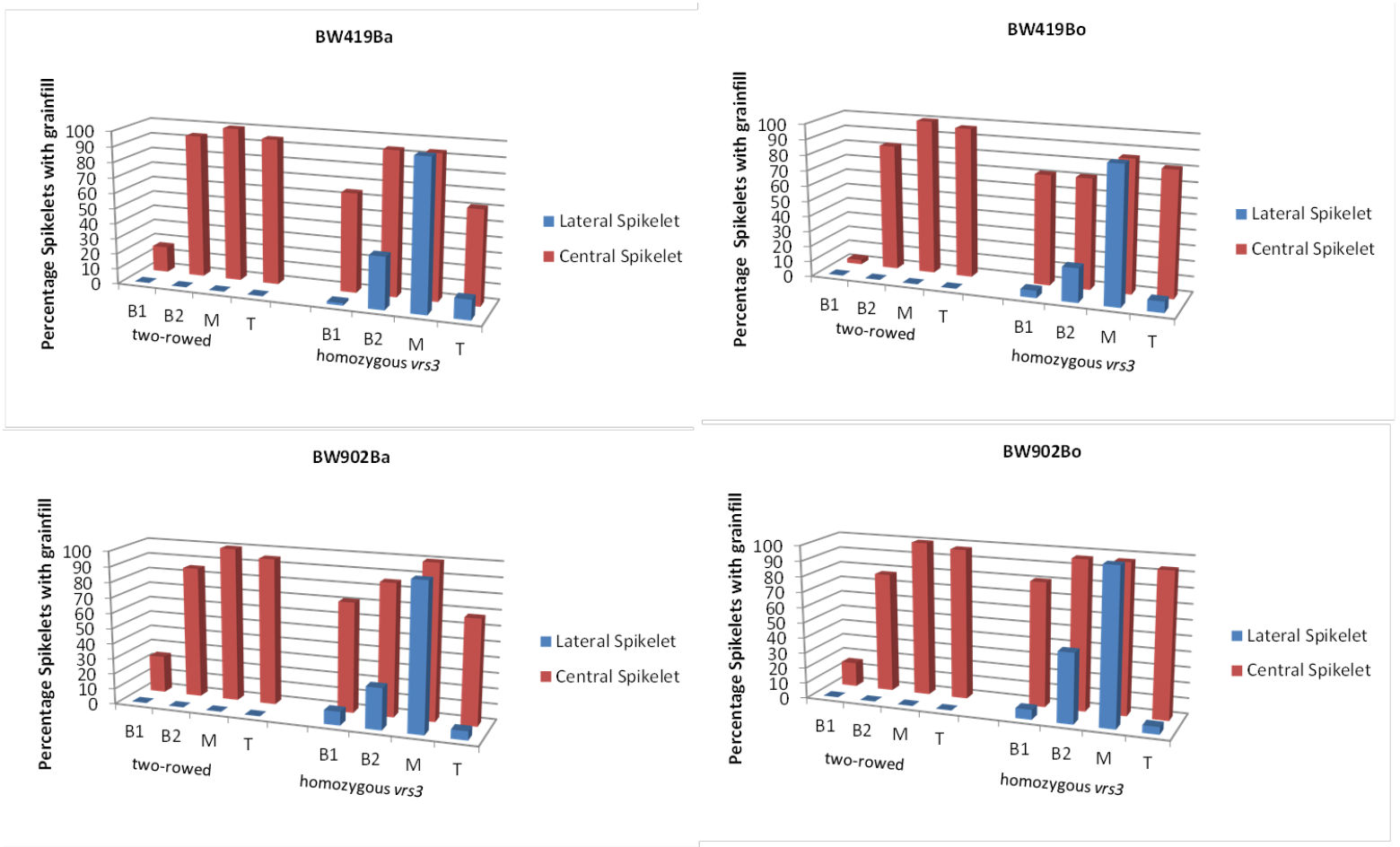


Figure 16: Comparison of spikelet grain-fill distribution across the four F_2 populations.

Individual graphs compare central and lateral spikelet grainfill in two-rowed and vrs3 homozygous spikes, across four positions of the spike B1= base spikelet 1, B2=base spikelet 2, M =middle spikelet, T=top spikelet.

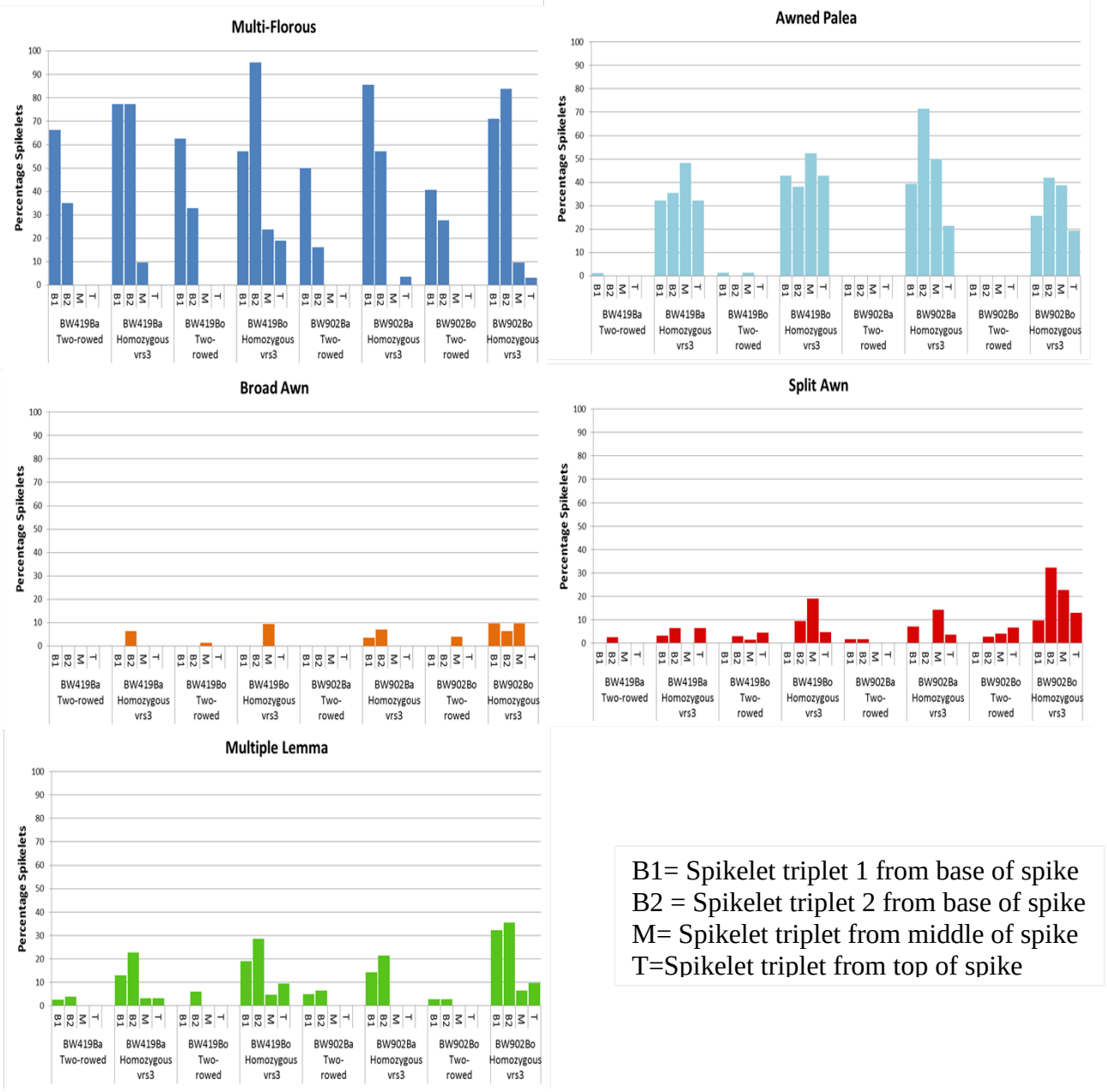


Figure 17: The distribution of inflorescence phenotypes across the four positions of the spike
Each bar represents the percentage of spikelet triplets sampled which showed the phenotype. Graphs show both two-rowed and vrs3 spikes separately.

The additional inflorescence morphological phenotypes, multi-florous, awned palea, broad awn, split awn and double lemma were observed at low frequency in the individual populations (Figure 17). To enable accurate assessment of any association between row-type and the presence of the phenotype, values were pooled according to row-type across the populations.

Multi-florous

The multi-florous phenotype was observed predominantly in the B1 and B2 positions of the spike. Despite being observed in the two-rowed spikes, test of association found a

significant association with the presence of the multi-florous phenotype and row-type across all four positions of the spike ($p<0.001$)with the strongest association at the B2 position and the weakest at the B1 position (Appendix 7).This would suggest that a proportion of the multi-floretted phenotype is associated with *vrs3* but some is also attributed to other background genetic factors.

Awned palea

The awned palea phenotype was observed at all positions across the homozygous *vrs3* spike. All positions had strong associations with the presence of the awned palea phenotype and row-type ($p<0.001$), with the strongest associations at the B2 positions (Appendix 7).

There were three observations of the awned palea phenotype in two-rowed spikes, all from crosses originating from BW419. Subsequent analysis of the segregation of these lines in the F_3 generation found them all to be heterozygous for *VRS3*. Before conclusions could be drawn as to the dominance relationship of *VRS3* and this phenotype a larger number of individuals within each population would have to be screened.

Broad awn

Broad Awns were observed more sporadically across the populations, with no observations at the T position of the spike. Significant associations with row-type were found for broad awns observed at the B1 ($p=0.008$) and B2 ($p<0.001$) positions, respectively with an excess of the phenotype observed in the *vrs3* homozygotes (Appendix 7).

However, no significant association was found between row-type and the broad awn phenotype at the M position of the spike. At this position, there were 5 observations of the broad awn phenotype in the homozygous *vrs3* spikes and 4 observations in the two-rowed spikes, with 3 of these confirmed as heterozygotes in the F_3 and the fourth of unknown genotype. The low number of observations means it is not possible to draw any conclusions as to whether these observations are due to background genetic effects or incomplete dominance of *vrs3*.

Split awn

The split awn phenotype occurred at all four positions across the spike, with the greatest prevalence observed in the BW902Bo population. Significant associations with row-type were found at the B1 ($p<0.003$). At the T position no significant association was observed, with equal numbers of observations in the two-rowed and the *vrs3* homozygous mutants (Appendix 7). At this position 7/8 of the observations in the two-rowed spikes were heterozygous for *vrs3* but investigations in larger populations would be required to ascertain any codominance relationships.

Multiple lemma

The multiple lemma phenotype was most prevalent in the Bowman crosses and showed a pre-dominance in the B1 and B2 positions of the spike. Analysis of associations with row-type found significant associations with the multiple lemma phenotype and row-type at all positions across the spike ($p<0.008$; Appendix 7).The occurrences of the double lemma phenotype in the two-rowed spikes were in both heterozygous *vrs3* and homozygous wild-type; this would suggest that not all occurrences of this phenotype can be attributed to *vrs3*, with other background genetic effects also influencing the phenotype.

It is clear from these results that the homozygous *vrs3* phenotype varies not only between different genetic backgrounds but also between individuals from the same genetic background. The phenotype was also found to vary significantly depending on the position of the spikelet within the spike, with those spikelets in the middle of the spike showing a more six-rowed mutant phenotype compared to those spikelets at the base of the spike.

2.3.1.6 *Pleiotropic phenotypes of vrs3*

To establish whether the *VRS3* locus influences traits in addition to spike morphology, data was recorded on spike emergence, tiller number, height, spike length and awn length in the Bowman and Barke F_2 populations.

Spike emergence

As a main effect row-type, did not have a significant effect on spike emergence, ($p=0.278$). However a significant interaction was detected between row-type and sowing date ($p=0.002$). Some plants were resown due to the initial sowing being eaten by mice; re-sown plants took an average 15 days less to reach heading compared to those sown initially, which most likely reflects the increasing temperature and day-length. In the earlier sown individuals no significant difference was detected between two-rowed or homozygous *vrs3*, however, within the later sown individuals, there was a significant difference in the heading date of two-rowed compared to *vrs3* mutant individuals ($p=0.002$), with two-rowed individuals heading two days earlier.

Parent cultivar also had a significant effect on the days to spike emergence ($p<0.001$) with F_2 individuals from the Bowman crosses reaching spike emergence on average 6 days earlier than Barke crosses; this difference persisted across sowing dates (Appendix 8).

Tiller number

Across the four populations tiller number varied markedly with a maximum of 118 tillers per plant and a minimum of 17. Inspection of residual plots found the data to be non-normal and a square-root transformation was applied prior to analysis to correct this. Row-type was found to have the largest significant effect ($p<0.001$); two-rowed individuals produced on average 54 tillers compared to a mean of 40 tillers of homozygous *vrs3* individuals (transformed means two-rowed 7.34, *vrs3* 6.35 SED (Standard error of the difference)= 0.083) .suggesting perhaps a compensatory effect for the additional energy associated with producing extra grain per node in the *vrs3* mutant individuals.

In general Bowman crosses produced significantly more ($p=0.016$) tillers than Barke, averaging 6 extra tillers per plant, but no significant difference with respect to sowing date. This effect was also seen with respect to the mutant allele in the cross; across both sowing dates BW419 crosses produced significantly more tillers than BW902

($p<0.001$), 55 tillers compared to 42 in the earlier sowing and 48 compared to 43 in the later sown individuals (Appendix 9).

Height to collar

The collar of the barley plant, marks the point of the plant at which the culm stops and the spike starts. The later sown plants, initially in a heated glasshouse environment, promoted a greater rate of stem extension in the plants, which led to a mean height 10 cm taller ($p<0.001$) than that of the earlier sown plants. This effect was independent of row-type, mutant allele or parent cultivar, which were all non-significant with regard to this trait.

Spike length

vrs3 mutant plants had significantly shorter spikes in comparison to two-rowed plants, 8.81 cm compared to 9.14 cm respectively (SED=0.10). In this case, as with tillering, it is potentially a compensatory effect, with potentially a reduced number of nodes of spikelet triplets produced per spike to account for the de-repressed out-growth of the lateral spikelets in the *vrs3* mutants. However , to determine if this is the case spikelet density would need to be calculated.

Significant differences in spike length were also detected with respect to parent cultivar, with spikes from the Barke populations on average 0.5 cm longer than those in the Bowman populations ($p<0.001$, SED=0.10) and with respect to Bowman Isoline, with BW902 population spikes 0.3 cm shorter than BW419 ($p<0.001$, SED=0.10) (Appendix 10).

2.3.2 *VRS3 mapping populations F_3 generation field trial phenotypes*

2.3.2.1 *VRS3 Barke and Bowman mapping populations F_3 segregation ratios*

To enable *VRS3* to be included as a co-dominant marker in the genetic mapping, the segregation of the *VRS3* phenotype was scored in the F_3 generation. This also enabled the assessment of the *vrs3* phenotype under field conditions. With respect to the crosses of BW419 and BW902 to Barke and Bowman, there was no evidence to suggest deviation from the expected 1: 2: 1 segregation ratio, supporting the hypothesis of a single recessive gene underlying the *VRS3* locus (**Table 5**).

Table 5: Segregation ratios of *VRS3* wild-type: mutant phenotype in the F_3 generation

Population	Total	Two-Rowed	Segregating	Six-rowed	Chi-Squared (2df)
BW419Ba	106	20	56	30	2.35 (p=0.308)
BW419Bo	88	23	43	22	0.07 (p = 0.967)
BW902Ba	89	17	44	28	2.73 (p=0.256)
BW902Bo	106	18	57	31	4.09 (p=0.129)

Spike emergence

In agreement with what was found in the polytunnel, row-type was found to have no significant effect on spike emergence but the parent cultivar in the cross did ($p<0.001$); Bowman crosses reached spike emergence significantly earlier ($p<0.001$) than Barke, 72 days after sowing compared to 78 days. The difference in days to reach spike emergence between Barke and Bowman is consistent between the polytunnel and the field, 6 days in both cases, suggesting the difference observed to be independent of environmental influences.

Unlike the comparable F_2 populations, there was no evidence of an effect of the allele of Bowman isoline present having a significant effect on days to heading.

Tillering

Tillering was scored on the populations at two-different time-points, prior to spike-emergence (Zadoks GS39) and post-harvest. At GS39, the flag leaf collar on the main-tillers of the plant is just visible, at this time-point the spikelets of the developing

inflorescence of the main tillers would be fully determined but secondary tillers are still being formed and therefore their developing inflorescene will be at different stages of differentiation. As with the F_2 plants, the cultivar in the cross was found to have a significant effect on tillering in the F_3, but in this case only at the post-harvest time-point ($p<0.001$) and not at GS39 ($p=0.563$), with Bowman producing on average 3 more tillers, 18.38 compared to 15.49, than Barke.

This effect was also seen with respect to row-type; no significant effect was observed at GS39 ($p=0.094$) but post-harvest a significant difference in tillering between two-rowed and homozygous *vrs3* was observed ($p=<0.001$) (Table 6).

Table 6: Mean tiller numbers at growth stage 39 and post-harvest for two-rowed and six-rowed in the Bowman and Barke F_3 populations
SED relate to transformed values. Values in brackets are untransformed tiller numbers.

	Two-rowed	Six-rowed	SED
GS39	3.030 (9.18)	2.979 (8.85)	0.0320
Post-Harvest	4.305 (18.53)	3.918 (15.35)	0.05094

The finding in differences in significance between the time-points suggests that at least two separate mechanisms are controlling tillering within the plant; primary tillering which in this case appeared to be unaffected by parent cultivar or genotype at the *VRS3* locus and secondary tillering, which was influenced by *VRS3* and other parent specific loci.

2.3.3 *Linkage mapping of VRS3*

The row-type phenotypes from F_2 and F_3 individuals from the BW419Ba, BW419Bo, BW902Ba and BW902Bo were used to establish a co-dominant marker for the *VRS3* locus. This was included with the F_2 BeadXpress SNP genotypes to map the position of the *VRS3* locus on chromosome 1H.

Table 7: Summary of the polymorphic markers used in linkage mapping

Population	Total Polymorphic SNP	Number of polymorphic SNP on chromosome 1H	Chromosomes to which other polymorphic SNP map
BW419Bo	10	7	4H
BW902Bo	6	3	7H
BW902Ba	170	10	All
BW419Ba	176	11	All

Table 7 summarises the polymorphic markers mapped in each population. Due to the near isogenic nature of both BW419 and BW902 to Bowman, the numbers of polymorphic SNP in the BW419Bo and BW902Bo populations were minimal. However, in both cases polymorphic SNP were present which according to the BOPA consensus map[148] are located on alternative chromosomes to the *VRS3* locus. This could be the result of residual heterozygosity or alternatively differences within the stock of Bowman used to make the Bowman isolines that could have introduced additional polymorphisms at random stages in the crossing cycle.

Following regression mapping individual markers contributed a mean chi-square <1, indicative of a good fit to the linkage map and no double recombinants were identified suggesting reliable marker map order. In the BW902Ba population, 11_10006 remained unlinked during the mapping of *VRS3*.

Due to a lack of polymorphic markers, in both the BW902Bo and BW419Bo populations the position of *VRS3* could not be mapped to an interval but only relative to the single flanking SNP marker, 9.2cM from 11_10597 (BW419Bo) and 3.3cM from 11_10833 (BW902Bo) (Figure 18a & b) . In the BW902Ba population *VRS3* mapped to a 55.7 cM interval between 11_10775 and 11_10438, this is in contrast to the BW902Bo population where 11_10438 mapped proximal to *VRS3* (Figure 18c). *VRS3* was mapped to the smallest interval, 37.2 cM, in the BW419Ba population, between markers 11_10833 and 11_10006 (Figure 18d).

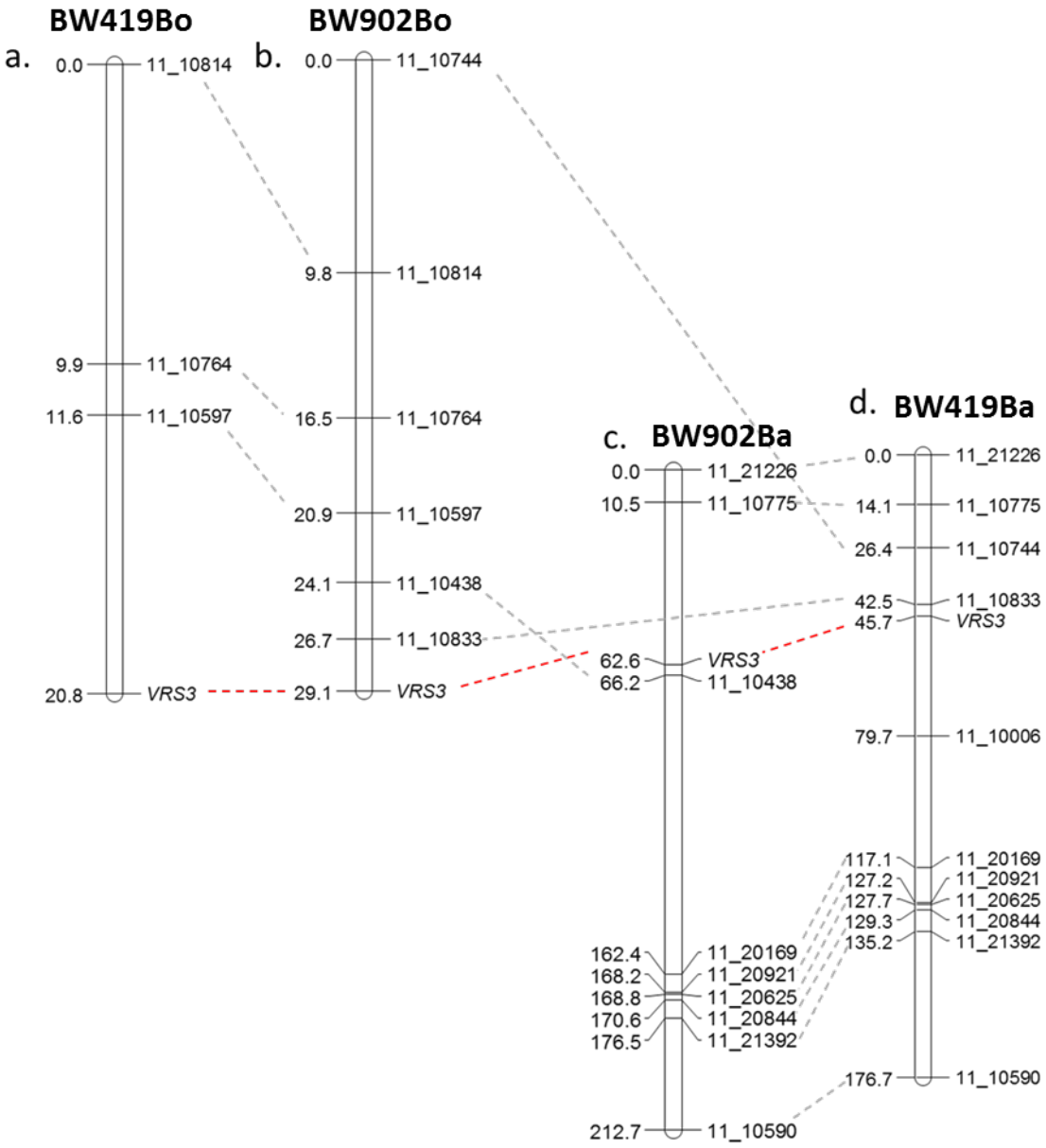


Figure 18: Linkage maps of the four F_2 populations with *VRS3* included as a co-dominant marker.
The dotted red lines link the differing positions of the VRS3 locus across the maps. The dotted grey lines link identical SNP across the maps. The maps refer to the following populations: a.BW419Bo; b.BW902Bo; c.BW902Ba; d.BW419Ba.

This first round of mapping localised *VRS3* to the same region identified from the genotyping of the Bowman Isolines by Druka *et al*[114].

The crosses to Barke proved the most informative but the lack of consensus on the position of marker 11_10438 meant an additional flanking marker needed to be identified to delimit the region of chromosome 1H within which *VRS3* mapped. Further

additional genotyping concentrated on the BW902Ba and BW419Ba populations due to their greater level of polymorphism.

SNP 12_30821 from the BOPA2 set of Illumina Golden Gate SNPs was identified from the BOPA 1 and 2 consensus map[65] as a possible flanking marker to *VRS3*, as it was located 14 cM from 11_10833. Pilot BOPA genotyping data[65] suggested it to be polymorphic between Barke and Bowman, and Barke and Bonus. Subsequent genotyping found 12_30821 to also be polymorphic between Barke and Hakata 2.

Genotyping 108 individuals from the BW419Ba and 89 individuals from the BW902Ba population with SNP 12_30821 substantially reduced the interval in which *VRS3* mapped to regions of 10.9 cM and 12.9 cM respectively (Figure 19). In addition, the revised mapping for BW902Ba altered the position of SNP 11_10438 relative to *VRS3* on the BW902Ba map and in concordance with its position in the BW902Bo map.

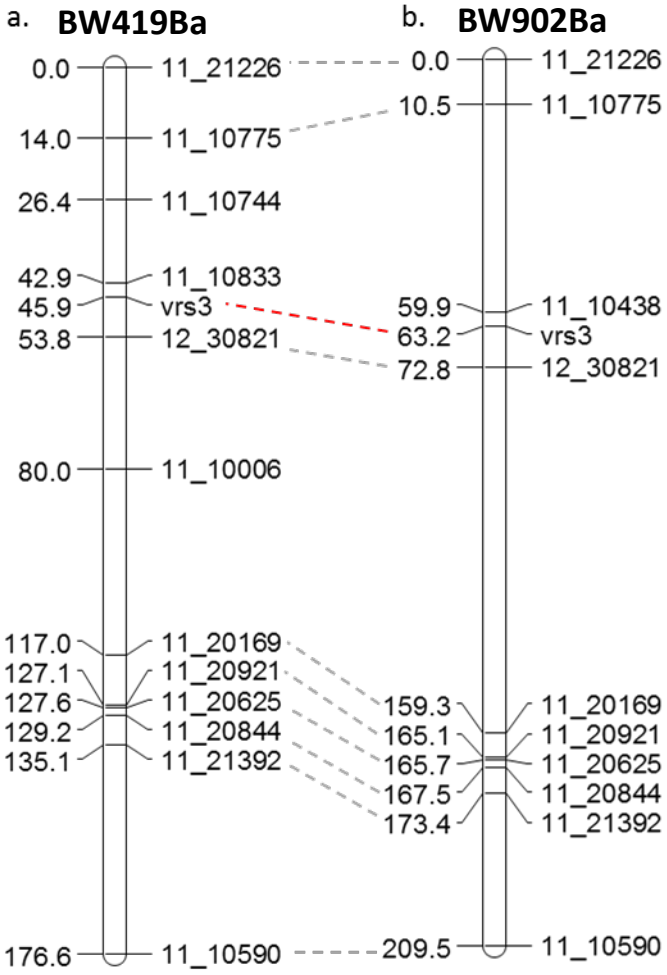


Figure 19: Revised genetic linkage maps for a)BW419Ba and b)BW902Ba.
Grey dashed lines link identical SNP across the two maps. The red dashed line links the position of VRS3 on the two maps.

2.3.4 ***Recombinant fine mapping***

Identification of flanking markers for *VRS3* provided the opportunity for fine mapping of the region to provide a small genomic segment for candidate gene studies. This required the identification of further polymorphic SNP markers with which to saturate the mapping interval.

From the consensus map[65], 36 BOPA1 and BOPA2 SNP either co-localised with or were positioned in the interval between 11_10438 and 12_30821. Using previously collected BOPA1 and BOPA2 genotyping information[65] potential polymorphic SNP in

the F_2 populations were predicted, resulting in the identification of 25 SNP for KASP assay design (Table 8). Whilst it would have been possible to re-genotype the populations with the entire BOPA2 set of SNPs with the Illumina Golden Gate platform, this was considered a far from a cost-effective option as most SNPs would be uninformative. Re-designing assays for the 25 SNPs on the KASP platform was the best option for quick and effective genotyping. Hakata 2, the cultivar from which the BW902 mutant originates has not been previously genotyped; therefore the prediction of likely polymorphic SNP in the BW902 populations was limited.

Table 8: SNP identified for use in fine mapping of *VRS3* and their predicted polymorphism prior to KASP design

The SNP delimiting the VRS3 mapping interval are highlighted in red. Those populations in which the SNP was known to be polymorphic are marked with a Y, those populations in which the SNP was known to be non-polymorphic are marked with an N, and those cases where it was unknown whether the SNP was polymorphic are marked with a U.

SNP	Chromosome	Consensus Map Position	Polymorphism	Population			
				BW902Bo	BW419Bo	BW902Ba	BW419Ba
12_30562	1H	74.4	A/G	U	N	U	N
12_31208	1H	74.4	C/G	U	N	U	N
11_10438	1H	74.4	A/G	Y	N	Y	N
11_21357	1H	76.96	C/T	U	Y	U	U
12_11169	1H	78.03	C/T	U	Y	U	Y
11_21000	1H	78.03	A/C	Y	Y	U	U
11_21312	1H	78.03	C/T	Y	U	U	U
12_30350	1H	78.03	C/T	U	Y	U	Y
12_30592	1H	78.03	T/G	U	N	U	Y
12_31134	1H	78.03	A/G	U	N	U	Y
12_31272	1H	78.03	T/C	U	N	U	Y
11_10833	1H	78.03	C/T	Y	N	N	Y
11_21217	1H	81.26	A/G	U	Y	U	U
12_30672	1H	81.26	G/T	U	Y	U	Y
12_30786	1H	81.26	C/G	U	N	U	Y
11_10520	1H	82.35	A/G	U	Y	U	U
11_20798	1H	82.35	A/T	U	Y	U	U
11_21361	1H	82.35	C/G	U	Y	U	U
12_30478	1H	82.35	A/G	U	Y	U	Y
12_30499	1H	82.35	A/G	U	Y	U	Y
12_30694	1H	82.35	A/G	U	Y	U	Y
12_30750	1H	82.35	A/G	U	Y	U	Y
12_30110	1H	82.35	A/G	U	Y	U	Y
12_30710	1H	83.42	A/G	U	N	U	N
12_10198	1H	85.57	C/T	U	N	U	N
12_30821	1H	88.33	C/G	U	N	U	Y
12_30753	1H	89.85	G/A	U	N	U	N
12_30304	1H	90.92	C/T	U	N	U	N

KASP genotyping of the parent lines found 2 of the 25 SNP to be polymorphic between BW902 and Barke, 12_30562 and 12_31208.

Informative recombination events were expected to be almost entirely confined to those lines that were recombinant between the flanking markers of the interval identified in the initial mapping. In both BW419Ba and BW902Ba this was 18 lines. These 36 lines

were genotyped with each of the SNP markers that were found to be polymorphic between their respective parents.

2.3.4.1 *BW902Ba recombinant genotyping and mapping*

Genotyping of recombinants from BW902Ba found no further recombination events between 11_10438 and the two additional BOPA2 SNPs, 12_30562 and 12_31208 so map resolution could not be improved (Figure 20). Therefore to further fine map in this population, Hakata 2 would have needed to be genotyped with other SNP markers to identify any further polymorphic SNP. Additionally more individuals from the original F_2 population could have been included in additional genotyping to increase the chances of identifying recombinants within the identified mapping interval.

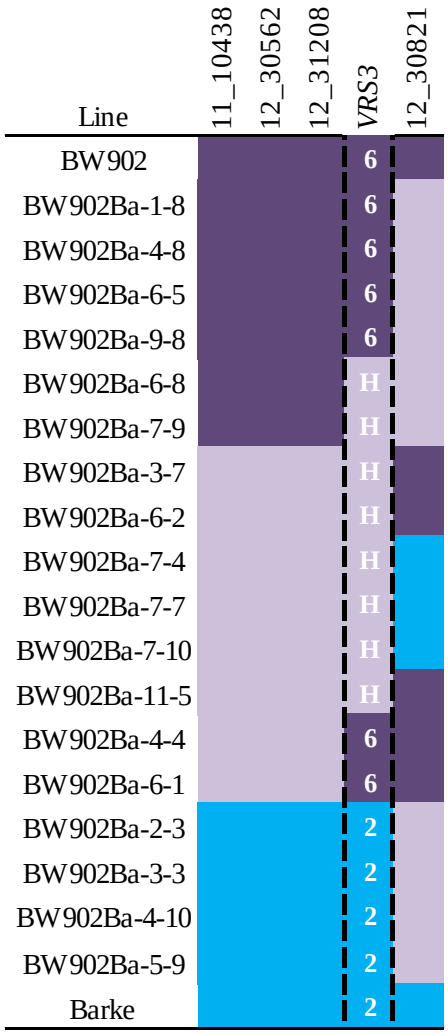


Figure 20: SNP haplotype across the 18 BW902Ba recombinant lines within the *VRS3* mapping interval

Indigo blocks represent BW902 alleles, lilac blocks represent heterozygous alleles and blue blocks represent Barke alleles. The individual F_3 phenotype at the VRS3 locus is shown in white text, 6—six-rowed, H —heterozygous, 2—two-rowed.

2.3.4.2 *BW419Ba recombinant genotyping and mapping*

By contrast, 15 SNP were polymorphic between BW419 and Barke and were therefore used to genotype the BW419Ba recombinants (Figure 21). Within recombinant BW419Ba-1-6 at marker 12_30672, the genotype recorded suggested a double recombination event. However, the probability of this double recombination event is $p<4.2x10^{-5}$; it is therefore more likely attributed to genotyping error and the genotype call was removed prior to linkage mapping.

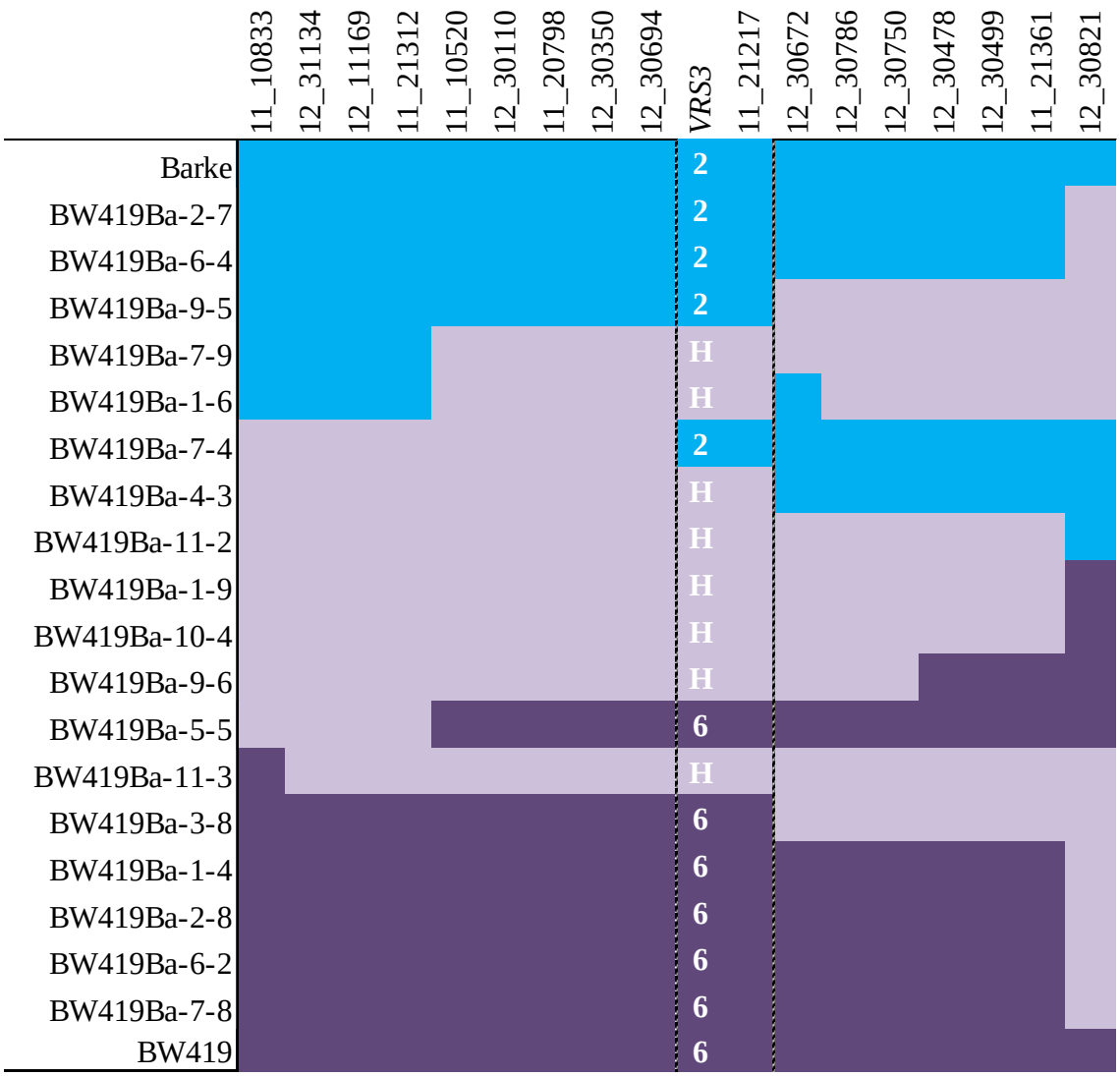


Figure 21: SNP haplotype blocks within the *VRS3* mapping interval delimited by flanking markers 11_10833 and 12_30821.
Indigo blocks represent BW419 alleles, lilac blocks represent heterozygous alleles and blue blocks represent Barke alleles. The individual F_3 phenotype at the *VRS3* locus is shown: 6 homozygous *vrs3*, H –heterozygous, 2-two-rowed. The 18 recombinants are listed on the left.

To ascertain estimates of recombination distances across the *VRS3* fine mapping interval, genotypes were imputed for 88 lines which in rough mapping had shown no evidence of a recombination event between 11_10833 and 12_30821. One line showed

evidence of a double recombination event between 11_10833 and 12_30821. As it is not clear where within the mapping interval the double recombination event is located, the genotypes for this line were not imputed but coded as missing values.

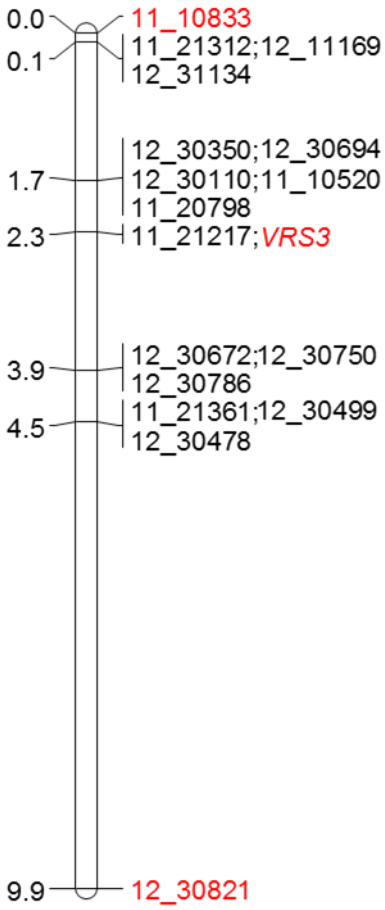


Figure 22: Genetic linkage map following fine mapping of *VRS3* in the BW419Ba population.
The flanking markers identified in the initial mapping and the VRS3 locus are all highlighted in red.

Regression mapping fitted all 18 markers within the first round of JoinMap. Individual marker contributions had mean chi-squared less than 1 and no significant double recombinants were identified indicative of reliable marker order.

Fine mapping refined the interval between flanking markers 11_10833 and 12_30821 by 1 cM, to a distance of 9.9 cM (Figure 22). The position of *VRS3* has been further refined to a 2.2 cM interval within which SNP 11_21217 co-segregated, indicative of tight linkage between the two loci. The interval was flanked by two blocks of markers in complete LD. Block one comprised 5 SNP, 11_10520, 12_30110, 11_20798, 12_30350 and 12_30694. The second block comprised 3 SNP: 12_30672, 12_30786 and 12_30750. Due to the lack of informative recombination events within these blocks

it is not possible to establish precise marker order within the blocks flanking *VRS3* from the fine mapping but they do delimit the region of the barley genome within which the gene underlying *VRS3* is most likely positioned.

2.3.4.3 *Identification of VRS3 candidate genes using conserved gene order with rice*

Due to the lack of available genome sequence data for barley when this experiment was conducted, the highly conserved gene order between barley and rice was used to further refine marker order in the region surrounding *VRS3* and to search for possible candidate genes.

Rice has a relatively small genome compared with barley which has facilitated more detailed genetic characterisation and full genome sequencing. As previously discussed in section 1.8.1, due to common ancestry, barley shares blocks of highly conserved gene order with rice.

Comparison of the *VRS3* fine mapping interval of chromosome 1H with rice using Strudel[146] , found it to have conserved synteny with the distal telomeric region of rice chromosome 10 (Figure 23).

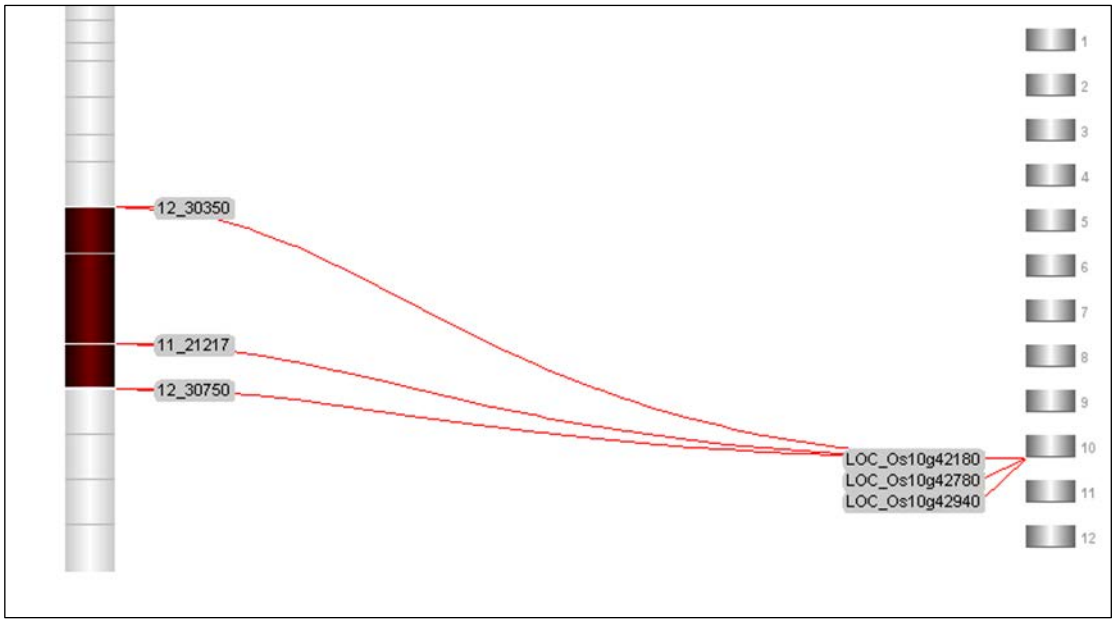


Figure 23: Conserved gene order between barley chromosome 1H and rice chromosome 10 in the *VRS3* mapping interval.
The region of the barley chromosome identified from fine mapping is shown on the left shaded in red, SNP markers from the flanking recombinant blocks, 12_30350 and 12_30750 are shown; as is 11_21217, the marker that co-segregated with the VRS3 locus. Their respective orthologous loci in rice are shown on the right.

To produce an initial list of possible candidate genes for *VRS3*, it was assumed that local conserved gene order was maintained between barley and rice. This assumption enabled the establishment of the predicted gene order within the recombinant blocks, identified in fine mapping of the BW419Ba population, through the order of their orthologous rice loci (Table 9).

Table 9: Establishment of the SNP marker order surrounding *VRS3* through the conservation of gene order with the rice physical map.

Barley SNP	Orthologous Rice Locus	Rice chromosome 10 position (bp)	
		Start	End
12_30350	LOC_Os10g42180	22685028	22682272
12_30110	LOC_Os10g42430	22857123	22859886
12_30694	LOC_Os10g42439	22875955	22861341
11_20798	LOC_Os10g42500	22920519	22923454
11_10520	LOC_Os10g42610	22965622	22964026
11_21217	LOC_Os10g42710	23039714	23035339
12_30786	LOC_Os10g42860	23113864	23111980
12_30672	LOC_Os10g42870	23114698	23117754
12_30750	LOC_Os10g42940	23146494	23148646

Assuming no localised genome rearrangements in barley, comparison with the rice physical map suggested the order shown in Table 9 for the markers within the recombination blocks flanking *VRS3*, with *VRS3* most likely to lie between 11_10520 and 12_30786 in barley. This related to a 148Kbp region in rice between loci LOC_Os10g42610 and LOC_Os10g42860.

2.3.5 *VRS3 candidate gene list*

All rice gene models predicted to lie within the interval defined by the rice equivalents of the flanking markers 11_10520 and 12_30786 were therefore considered potential candidates for *VRS3* (Table 10). Whilst any one of the 16 genes remained candidates for *VRS3*, it was most likely that one of the loci closest to 11_21217 was the candidate for *VRS3*, given their co-segregation in the linkage mapping of *VRS3* in the BW419Ba population or alternatively a non-syntenic gene.

Table 10: Potential candidate genes for *VRS3*

Barley SNP	Rice Locus	Rice Chromosome 10 position (bp)		Putative Gene Function
		Start	End	
11_10520	LOC_Os10g42610.1	22965622	22964026	expressed protein
	LOC_Os10g42620.1	22981601	22986324	dihydroflavonol-4-reductase
	LOC_Os10g42630.1	22989900	22986842	expressed protein
	LOC_Os10g42660.1	23007668	23005547	expressed protein
	LOC_Os10g42670.1	23011807	23008540	glycosyl hydrolases family 16
	LOC_Os10g42690.1	23025714	23031771	jmjC domain containing protein
	LOC_Os10g42700.1	23032206	23035098	CS domain containing protein
11_21217	LOC_Os10g42710.1	23039714	23035339	RCD1
	LOC_Os10g42720.1	23044693	23050740	acyltransferase
	LOC_Os10g42724.1	23052138	23055070	VHS and GAT domain containing protein
	LOC_Os10g42750.1	23062454	23066292	CSLD1 - cellulose synthase-like family D
	LOC_Os10g42780.1	23079075	23075396	lrgB-like family protein
	LOC_Os10g42790.1	23082297	23079444	cytidylyltransferase domain containing protein
	LOC_Os10g42800.1	23086585	23083220	AMP-binding enzyme
	LOC_Os10g42820.1	23093419	23088097	early-responsive to dehydration protein-related
	LOC_Os10g42830.1	23099812	23095271	transporter family protein
	LOC_Os10g42840.1	23103301	23100365	NADH dehydrogenase 1 alpha subcomplex subunit 12
12_30786	LOC_Os10g42860.1	23113864	23111980	ribosomal protein L35

2.4 *Discussion*

Genome-wide SNP analysis in the F_2 mapping populations BW419Ba, BW419Bo, BW902Ba and BW902Bo confirmed the location of *VRS3* on chromosome 1H, consistent with previous studies[89,114]

Further targeted marker genotyping positioned *VRS3* in a 2.2 cM region between markers 11_10520 and 12_30786, suggesting that *VRS3* lies just outside the region previously delimited by the polymorphic markers in the BOPA1 genotyping of BW419 and Bowman[114]. The most likely reason for this is the apparent fore-shortening of the BW419 introgressed segment due to lack of polymorphism between Bonus and Bowman in the region proximal to 11_10520.

Comparison of the fine mapping interval in barley with the rice physical map has identified 16 candidate loci within the 2.2 cM region, this is relatively few candidates given previous estimates of 4600-5800 genes over 140 cM [149] for chromosome 1H. Gene density is known to vary greatly across the chromosome[150] suggesting *VRS3* may lie in a region of relatively low gene density. Another reason underlying this difference is the interval on the rice physical map was taken as the shortest physical distance between markers from the flanking blocks of recombination, i.e. 11_10520 and 12_30786. This is 1/3 of the physical distance compared to the markers at the same respective positions on the *VRS3* genetic map but, separated by the greatest distance on the rice physical map, 12_30350 and 12_30750. However, the co-segregation of *VRS3* with 11_21217 gave confidence in this approach.

Through comparative re-sequencing of the most promising candidates within the *vrs3* Bowman isolines, *vrs3* induced mutant allelic series, and respective parental germplasm identification of the candidate locus underlying *VRS3* was possible (Chapter 3).

2.4.1 *The vrs3 phenotype*

The phenotypes recorded in the F_2 mapping populations did not differ markedly from the published *vrs3* phenotypes[135,136], with lateral grain-fill and awn-development showing progressive development up the spike and the awned palea phenotype observed in a number of cases. However, in addition to the previously reported phenotypes, multi-florous, broad awned, split awn and multiple lemma phenotype were also observed. These observations were predominantly in homozygous *vrs3* mutant spikes, but they were not observed in all *vrs3* spikes, suggesting their occurrence is due to differing background genetic factors e.g. factors controlling the timing of cell

differentiation within the developing spike or environmental factors e.g. temperature during spike differentiation.

Although the majority of observations of these additional phenotypes were in the *vrs3* homozygous individuals, there were observations of the phenotype in two-rowed spikes, the majority of which were heterozygous. From the low number of observations, it is not clear whether these are the result of partial incomplete dominance of *vrs3* or due to background genetic factors. To reliably ascertain this, screening of sufficiently large populations, thousands of individuals, would be required.

Analysis of spikes from the Bowman and Barke F_2 mapping populations suggests the BW902 mutant to produce a stronger six-rowed phenotype, with enhanced lateral grain-fill and awn development compared with BW419. Once the gene underlying *VRS3* is identified, it will be possible to see if there is any correlation in the position or type of mutation present in the Bowman isolines and the strength of the phenotype observed. There was also evidence that the phenotypes observed depended on the cultivar background into which they were crossed. Those crossed into a Bowman background showed a stronger phenotype with increased lateral awn development and lateral grain-fill compared to the same cross to Barke suggesting interaction between the allele of *vrs3* and background genetic loci.

The irregularity of some aspects of the *vrs3* phenotypes for example the presence/absence of awned palea, or additional spikelets, suggests that expression of the *vrs3* locus is likely to be environmentally sensitive. Changes in spike morphology with respect to environment have previously been found in the semi-dominant barley hooded mutant. Hooded mutants rather than producing a lemma awn, produce an additional floret structure or hood. When exposed to extreme cold temperatures at the hood cushion development stage, spikes reverted to producing awn-like structures rather than hoods [151]. Environmental sensitivity at intermedium loci has also previously been noted in the *vrs4* mutants with a more severe phenotype observed in *vrs4.l* mutants sown in spring compared to autumn.[130]

Within wheat, the formation of supernumerary spikelets has been linked to the photoperiod sensitivity gene, *Ppd-1;* under short-day conditions, day-length sensitive wheat was found to produce a higher frequency of super-numerary spikelets compared to day-length insensitive wheat[152]. This does not explain the findings of supernumerary spikelets in this study as all mapping population individuals have a photoperiod

insensitive (*ppdH-1)* spring growth habit[153]. However, the study in wheat also linked the supernumerary spikelets to reduced levels of expression of the floral meristem promoter gene *FT*, independent of photoperiod; this could potentially underly the supernumerary spikelets observed in the mapping population but further expression studies would be required to confirm this.

2.4.2 ***Pleiotropic phenotypes***

Tillering

Across all populations, row-type was found to have a significant effect on tiller number, with two-rowed plants producing a greater number of tillers than *vrs3* mutant or six-rowed. However, results from the F_3 field trial suggest this difference occurs after GS39.

Barley tillers arise from the shoot apical meristem (SAM); after formation the SAM can lie dormant until signalled for outgrowth[20]. Primary tillers branch from SAM on the main stem, with secondary tillers branched from primary tillers and tertiary tillers branched from secondary tillers[154]. The signals that lead to bud outgrowth are still largely unknown in barley, but in other species auxin, cytokinins, gibberellins and strigolactones have been implicated[155]. Environmental cues including nitrogen availability, plant density and temperature have been shown to influence tiller number [25,27].

Consistent with the observation of a significant difference in tiller number between row-types occurring after GS39, an earlier comparative study of two-rowed barley cultivar Proctor with six-rowed barley cultivar Clermont found six-rowed barley Clermont to produce fewer tillers, primarily due to the lack of production of tertiary tillers. This study also found Clermont to reach maximum spikelet primordia differentiation sooner than Proctor. A mechanism of tiller outgrowth repression at the point when the main culm reached maximum spikelet primordia differentiation was proposed to account for the difference in tillering capacity[156]. However, this may not explain the findings observed in this study as although not measured, it would be expected that maximum spikelet primordia in spikes from the main tillers would occur prior to growth stage 39.

Other studies have assigned the differences in tiller number between two-rowed and six-rowed to increased tiller mortality in six-rowed, possibly due to competition for available nutrient to support increased spikelet production in six-rowed plants. [157] The

difference between the Bowman and the Barke crosses from the F_2 and F_3 phenotypes may also be due to differences in tiller mortality rates, in the genotype registration document for Bowman it is noted that Bowman has a low level of tiller abortion[115].

Barley tillering mutants are also associated with changes in inflorescence architecture, with the *low number of tillers 1 (lnt1)* mutant, found to have occasional missing spikelets in the middle and upper portions of the spike[129]. In addition to significantly fewer tillers the *absent lower laterals 1 (als1)* mutant is characterised by uneven distribution of spikelets across the spike, and absent lower lateral spikelets[158], suggesting a complex interplay between spike and tiller development within which *VRS3, INT-C and VRS1* play an as yet undetermined role.

Spike emergence

Whilst the *VRS3* locus was found not to affect spike emergence, the cultivar in the original cross did, with Bowman crosses reaching spike emergence six days prior to Barke. Spike emergence is underlain by a complex genetic network which responds to a number of environmental cues; some, for example *Ppd-H1* and *Ppd-H2*, are associated with day-length cues[153] whereas other loci, e.g. *eam-8 and eps2* are daylength insensitive[66,159]. Selection for alleles at these loci is largely dependent on the environment in which the original cultivar was bred; for example allele 3 of *eps2* is fixed in Northern European spring barley cultivars as the late flowering phenotype is thought to be advantageous for prolonged grain-fill and ripening, under cool growth conditions[66].

In the case of Bowman and Barke, both contain the recessive *ppd-H1* allele at the locus[153,159] and therefore it cannot be segregation at this locus underlying the difference in spike-emergence. Bowman, originating from North America, and Barke, originating from Germany, possess different alleles of *HvCen*, at the *eps2* locus, with Bowman allele I and Barke allele III[66]. The same allelic difference is observed in Morex and Steptoe (JHI unpublished), with bi-parental mapping populations finding the Steptoe allele at this locus results in later spike emergence[160], therefore the difference in heading observed between the Bowman and Barke populations is most likely to due to allelic difference at the *HvCen* locus. In addition Bowman and Barke segregate at SNP 12_30096 on chromosome 3H (JHI unpublished genotyping), associated with the semi-dwarfing locus *SDW1*[161]. In addition to its beneficial reduction in height, the *sdw1* phenotype also features a prostrate growth habit and delayed spike emergence[12,162,163], suggesting segregation at this locus could also be a factor in delayed spike emergence observed in Barke.

BW419 and BW902 gave a significant difference in spike-emergence in the F_2 generation. However, this difference was not observed in the F_3 generation. Whilst it could be an environmental effect, with the warmer conditions in the polytunnel promoting the difference between the two, it could equally be due to spatial effects within the polytunnel. F_2 populations were sown as non-randomized blocks of crosses, therefore any spatial effect influencing spike emergence would be confounded with the cross; these effects would be lost in the field as a replicated, randomized design was used.

Spike length

The significantly shorter spike length in the homozygous *vrs3* compared to the two-rowed spikes suggests that either the rachis-internode length is being affected by the *VRS3* locus or the number of spikelets per spike. Without counting the number of nodes, it will not be possible to identify which case is true. If it is the number of spikelets per spike, maximum spikelet primordia are determined at the awn primordia stage of plant development[164]; this is followed by abortive development of unwanted spikelets between the spikelet primordia stage and the emergence of the awn tips from the boot.[164] This would suggest two possibilities for the difference observed between two-rowed and six-rowed spikes, either fewer spikelet primordia are differentiated initially or a greater number undergo abortive development prior to anthesis. Previous work has suggested that six-rowed spikes abort a greater number of spikelets than two-rowed[164], with the greatest levels of abortion at the top of the spike[165]; a number of reasons have been proposed for this including the competition from other spikelets for assimilates[166] and differences in development of vascular tissue[165,167].

The significant difference in spike length observed between Bowman isoline alleles in the F_2 Bowman and Barke populations could again be confounded with the spatial effects associated with heading, alternatively the spike phenotype of Hakata 2 is characterised by a dense spike phenotype; if the locus which underlies this is retained within the introgressed segment of BW902 then the dense spike phenotype would persist despite the largely Bowman genetic background.

The difference observed in spike length between Barke and Bowman could be related to the later spike emergence associated with Barke. Previously a rapid rate of stem extension has been associated with a higher level of spikelet abortion[168], possibly due to competition for nutrients[166] required for the two distinct developmental functions; this would explain the longer spikes observed in the Barke compared to Bowman populations, as Bowman populations reached spike emergence significantly earlier than Barke.

2.4.3 ***Mapping VRS3***

The BW419Ba population was the most informative of the six F_2 populations, primarily due to the greater level of polymorphism between Bowman/Bonus and Barke and the greater amount of previously collected genotypic information on these cultivars. The greater levels of polymorphism in the regions flanking the *VRS3* locus in the BW419Ba

population helped to delimit the region in relative close proximity to *VRS3* unlike the BW902 population in which no polymorphic SNP were identified in the intervening 7.6 cM region between the *VRS3* locus and 12_30821.

Using the F_2 crosses to Bowman proved problematic for mapping due to their highly monomorphic nature. If an alternative strategy of bulk segregant analysis combined with next-generation sequencing approaches had been employed these populations may have proved more informative. Utilising this approach, separate pools of DNA from two-rowed *VRS3* individuals and homozygous six-rowed *vrs3* individuals would be subjected to reduction in genome complexity via exome capture and then sequenced using next-generation sequencing approaches,. Subsequent analysis of the sequence data should identify alleles which segregate 3:1 with the wild-type:mutant phenotype and map to the interval of chromosome 1H previously identified enabling the identification of candidate loci for *VRS3*. This technique has been successfully employed in the identification of the gene underlying the barley multi-noded dwarf phenotype [169]. A potential disadvantage of this approach is whilst the exome capture array targets a large portion of the barley exome, it does not capture all of it, and therefore there is the possibility that the sequence of interest would not be captured [170].

An alternative more targeted approach would have been to use RNA seq of Bowman, Barke, Bonus, Hakata 2 and the Bowman isolines to identify SNP between the lines. Any SNP identified within the loosely defined interval from the BOPA1 genotyping of the Bowman isolines could subsequently be used to create either a custom XOPA[171] to multiplex genotype individuals from the F_2 populations or alternatively allele-specific PCR assays with the subsequent genotypic information used for linkage mapping.

3. Identification and DNA Characterisation of *VRS3*

3.1 *Introduction*

In spite of marked differences in floral morphology, a number of conserved genes and mechanisms have been identified between cereal species especially with respect to floral development. For example, the *tassel-sheath 1*[172] mutant of maize, *neck-leaf1*[173] mutant of rice and *third outer glume 1*[174] mutant of barley are all characterised by the formation of leaf-like bracts in their respective tassels, panicles and spikes. In all cases, mutation of a conserved orthologous GATA domain containing transcription factor was found to be responsible for the loss of bract suppression [79,175]. A similar situation arises with respect to the inflorescence mutants: rice *frizzy panicle 1*[176], maize *branched silkless1*[177], barley *compositum 2*[178] and 'miracle' wheat, all characterised by a highly branched phenotype. Independent studies have found them all to be due to mutation of orthologous *ERF* transcription factors[179-181], which suppress branch formation within the spikelet meristem.

Additionally, despite the row-type phenomenon being specific to barley, of the previously identified row-type loci *INT-C*[72]and *VRS4*[130] both have orthologues in rice and maize (Sections 1.9.4.2 & 1.9.4.4).

Vrs3 has been previously localised to sixteen candidate genes, in a 2.2 cM mapping interval, on barley chromosome 1H (Chapter 2).

Given the large number of orthologues between cereals species with respect to inflorescence architecture, a candidate gene approach, utilising prior knowledge of the function, cellular localisation and mutant phenotypes of the *Vrs3* candidates in other species was taken to establish an order of priority in which to investigate the sixteen candidate *Vrs3* loci.

3.1.1 ***The vrs3 induced mutant allelic series***

In addition to the two *vrs3* mutant alleles *int-a*.1 (BW419) and *vrs3*.f (BW902), introgressed within the Bowman near isogenic line collection, a further 30 morphological mutants, were classified as alleles of the *VRS3* locus through reciprocal crossing of intermedium mutants at the Svalöv Institute, Sweden[136]. The *vrs3* mutant accessions are maintained within the NordGen genetic stock centre, with detailed records of their parental cultivars and mutagens kept. Although mutants of the *VRS3*

locus, these induced mutants were classified under the intermedium nomenclature and are therefore designated *intermedium-a (int-a).*

The type of mutagen with which the mutant alleles were induced varies amongst the allelic series, with earlier mutants produced by X-ray and other forms of ionising radiation and later alleles from ethylene derived chemical mutagens[182]. In addition to the type of mutagen, the cultivar within which the mutations were induced also varied, with *int-a* mutants induced in the spring barley cultivars Bonus, Foma, Kristina and Hege. Table 11 summarises the *int-a* allelic series, the mutagens and cultivars underlying each of the accessions.

Table 11: The *vrs3* induced mutant allelic series

NGB accession numbers refer to accessions within the Nordic Genetic Resource Center (NordGen)[93]; GSHO accession numbers refer to accessions within the Germplasm Resources Information Network database (GRIN), USDA[183].

Genebank Accession	Allele	Cultivar Background	Mutagen	Year induced
NGB115419	*int-a.*1	Bonus	X-rays	1943
NGB115420	*int-a.*2	Bonus	radon	1951
NGB115426	*int-a.*8	Bonus	X-rays	1957
NGB115427	*int-a.*9	Bonus	EI	1958
NGB115428	*int-a.*10	Bonus	EI	1958
NGB115432	*int-a.*14	Foma	DES	1959
NGB115435	*int-a.*17	Foma	EI	1960
NGB115439	*int-a.*21	Foma	EMS	1961
NGB115445	*int-a.*27	Foma	n-PMS	1963
NGB115448	*int-a.*30	Foma	n-PMS	1963
NGB115449	*int-a.*31	Foma	neutrons neutrons EMS neutrons	1963
NGB115450	*int-a.*32	Foma	EMS	1963
NGB115452	*int-a.*34	Foma	EI	1964
NGB115453	*int-a.*35	Foma	PDSADEE	1964
NGB115455	*int-a.*37	Foma	neutrons	1966
NGB115464	*int-a.*46	Kristina	neutrons	1966
NGB115469	*int-a.*51	Kristina	IPMS	1971
NGB115470	*int-a.*52	Kristina	X-rays	1972
NGB115472	*int-a.*54	Kristina	EMS	1972
NGB115473	*int-a.*55	Kristina	IPMS	1972
NGB115477	*int-a.*59	Kristina	SA	1977
NGB115479	*int-a.*61	Kristina	EMS	1974
NGB115482	*int-a.*64	Bonus	IPMS	1976
NGB115489	*int-a.*71	Bonus	SA	1976
NGB115492	*int-a.*74	Bonus	EMS	1978
NGB115495	*int-a.*77	Bonus	EMS	1978
NGB115497	*int-a.*79	Bonus	SA	1979
NGB115504	*int-a.*86	Bonus	SA	1979
NGB115506	*int-a.*88	Bonus	SA	1979
NGB115520	*int-a.*102	Hege	SA	1980
NGB115521	*int-a.*103	Hege	SA	1980
GSHO 774	*vrs3*.f	Hakata 2	Gamma-ray	-

EI=Ethylene Imine, DES= diethyl sulfate, EMS= ethyl methanesulfonate, n-PMS = n-propyl methanesulfonate, PDSADEE=propanedisulfonic acid diethylester, IPMS= iso-propyl methanesulfonate, SA= Sodium azide

This collection of *vrs3* alleles represents a valuable means of confirming the successful identification of the gene underlying *VRS3*. If correctly identified, it would be expected that re-sequencing of the candidate gene from genomic DNA isolated from these accessions would identify polymorphisms when DNA sequence of the candidate gene is aligned with that from the respective two-rowed cultivar of origin. This means of validation has been used successfully in the identification of the genes underlying other traits including the row-type loci *INT-C* and *VRS1,* and the *EARLINESS PER SE 2 (EPS2)* locus.[66,72,78].

3.1.2 ***VRS3 diversity within cultivated germplasm***

GWAS (section 1.8.4.2) have previously been used to investigate associations between row-type and regions of the barley genome across a number of different cultivated barley germplasm collections. In particular, this technique was used in the successful identification of the gene underlying *INT-C*. The panel used in the identification of *INT-C* comprised 190 cultivars of Western European and North American origin genotyped with 2463 genomewide SNP. Nine significant loci were found across 6 chromosomes; including one on chromosome 4H in the region of *INT-C*. The most highly associated SNP, 11_20606 was found to be eight rice gene models, from the underlying TCP gene. Despite identifying *INT-C* and finding a significant association with the previously identified *VRS1* locus on chromosome 2H. No significant association with row-type in the vicinity of the *VRS3* locus was found[72]. A similar result was found in a GWAS of 500 U.K. elite barley lines genotyped with 1536 SNP; associations which co-located with regions corresponding to *INT-C* and *VRS1* were identified but no significant associations were found on the short arm of chromosome 1H around the *VRS3* locus[184].

In contrast, GWAS of more globally diverse barley association panels has suggested some significant associations with row-type in the region of *VRS3*. In a study of 1860 barley accessions from the NGSC BarleyCore collection[185], including both cultivated and landrace accessions, genotyped with 6224 iSelect[66] SNPs, a significant association with row-type was found in the vicinity of the *VRS3* locus[67]. The most highly associated SNP was 11_10933 which maps <0.5 cM (IBGSC map) from 11_21217, the SNP found to co-segregate with *VRS3* in the bi-parental mapping (Chapter 2). A similar study focussed on a panel of 224 globally diverse, cultivated and landrace spring barleys genotyped with 957 SNP. This study identified 3 regions of barley chromosome 1H associated with row-type, one of which was within 1 cM of the *VRS3* locus[186] indicative

of possible natural variation with respect to *Vrs3* between two-rowed and six-rowed cultivars.

3.1.2.1 *Chapter summary*

I have combined reference genome resources and the *vrs3* induced mutant allelic series to confirm the identity of *VRS3* as a gene encoding a JMJC histone demethylase. Subsequent investigation of sequence diversity within a panel of diverse cultivated germplasm identified three allelic variants of *Vrs3*. Further investigation of interspecific diversity established that *VRS3* is highly conserved across a diverse range of plant species.

3.2 *Materials and methods*

3.2.1 ***Orthologous sequence identification***

3.2.1.1 *Rice —LOC_Os10g42690*

Genomic and FlcDNA sequence for the priority candidate of *Vrs3*, LOC_Os10g42690 were obtained from the Rice Genome Annotation Project [147].

3.2.1.2 *Barley genomic sequence—Morex*

BLAST of the LOC_Os10g42690 genomic DNA sequence against the Morex genomic reference sequence (2011 assembly) identified the 8,413 base pair (bp) contig 5669 as the best match. This contig had putatively been assigned to chromosome 1H, giving confidence that this was the orthologue of LOC_Os10g42690 (bit score=347, E=3e-92). Reciprocal BLAST of contig 5669 against the rice genome annotation project database identified LOC_Os10g42690 as the top hit (E=4e-93, bit score=175).

3.2.1.3 *Barley genomic sequence—Bowman*

BLAST of the LOC_Os10g42690 genomic DNA sequence against Barley Bowman Assembly 23 Oct 2010[57] identified the 8,499bp contig 845567 as the best match. (bit score=347, E=3e-92). Reciprocal BLAST of contig 845567 against the rice genome annotation project database identified LOC_Os10g42690 as the top hit (E=4e-93, bit score=175).

3.2.1.4 *Barley flcDNA and predicted protein sequence—Haruna Nijo*

BLAST of the Genbank DNA database[187], with Morex contig 5669 and Bowman contig 845567, identified the 3444bp FlcDNA clone AK368672 as the best match (99% identity, E=0.0, bit score=1683). Reciprocal BLAST of AK368672 against both the Bowman Assembly 23 Oct 2010 and Morex 2011 assembly identified contigs 5669 and 845567 as the top hits respectively.

3.2.1.5 *Sequence alignment*

All sequence alignment was carried out within Geneious 6.1(Biomatters Ltd). Sequence alignments used the MUSCLE alignment algorithm[188], with default parameters.

3.2.1.6 *Functional domain prediction*

InterPro Scan[189] was used to predict the functional domains within the peptide sequence predicted from FlcDNA AK368672.

3.2.2 ***Candidate gene re-sequencing across the vrs3 allelic series***

3.2.2.1 *Barley germplasm*

All *int-a* accessions from the *vrs3* allelic series were sourced from NordGen (The Nordic Genetic Resource centre). Hakata 2, the germplasm in which the *vrs3*.f mutation was induced, was sourced from the NIAS genebank, Japan, accession number, JP15419. The *vrs3*.f allele was sourced from NPGS (National Plant Germplasm System), USA accession number GSHO 774. All other parental germplasm was sourced from JHI (James Hutton Institute) stocks.

3.2.2.2 *Tissue harvest and DNA extraction*

Barley grains from the *vrs3* allelic series and parent cultivars were germinated in Petri-dishes lined with filter paper. Petri dishes were watered with 3 ml of distilled water prior to being sealed with parafilm. After a week at room-temperature, 100 mg of fresh leaf tissue was harvested per grain. Tissue was immediately frozen in liquid nitrogen.

Prior to DNA extraction, frozen leaf tissue was ground to a fine powder using a micro-pestle. DNA extractions were then carried out using the QIAGEN DNeasy plant minikit as per the manufacturer's protocol (July 2006 version) (http://www.qiagen.com/gb/products/catalog/sample-technologies/dna-sample-technologies/genomic-dna/dneasy-plant-mini-kit/#resources). Genomic DNA was eluted into a total volume of 50 µl with the final elution in two stages. DNA eluent from the initial stage was passed back through the column for a second time to increase final DNA concentration.

DNA quality was ascertained by gel electrophoresis, using a 1.5% agarose gel stained with SYBR-Safe (Invitrogen) DNA gel stain.

3.2.2.3 *Primer design for amplification of candidate VRS3 locus*

Overlapping primer pairs were designed to the Morex genomic DNA sequence. 16 primer pairs were designed to span the *VRS3* region. Primers were designed using Primer3 design software[190] with the following design parameters: primer size -21 nucleotides, maximum self-complementarity 6, maximum 3' self-complementarity 2.00, maximum length of polynucleotide string 4.00, Number of CG clamp -2. Appendix 12 lists the primer sequences and their relative positions; each primer pair was designed to span between 500-700 nucleotides.

Primers were BLASTed against the NCBI nucleotide database[191] to establish unique sequence.

3.2.2.4 *PCR and Sanger sequencing conditions*

All PCR amplification, prior to sequencing, used a 10 µl reaction volume. Each PCR reaction contained 50 ng genomic DNA, 0.2 mM dNTPs, 0.5 units QIAGEN HotStarTaq DNA Polymerase, 1 x PCR buffer and 1 µM Forward and Reverse primers (Sigma-Aldrich). Optimised primer conditions were established for each primer pair (Appendix 13). In preparation for sequencing, unincorporated primers and nucleotides were removed from the PCR reaction using Illustra™ ExoStar™ 1-Step cleanup reaction following the manufacturer's guidelines.

Sequencing reactions used Big Dye v3.1; approximately 15 ng of cleaned PCR template was included in each reaction. Cycling conditions were as follows: 96°C for 1 min; 25 cycles of 96°C for 10 sec, 50°C for 5 sec, 60°C for 4 min. Prior to analysis on an ABI3730 capillary sequencer reactions were purified using EDTA/ethanol precipitation.

3.2.3 ***BeadXpress genotyping of induced mutants***

Using the method previously described in section 2.2.6, *int-a* alleles and the two-rowed cultivars in which the mutations were induced were genotyped using the BeadXpress 384 genomewide SNP genotyping platform[142].

3.2.4 ***Dendrogram creation***

Dendrograms were created using the Cluster Analysis function within PAST 2.17[192]. The Hamming simple matching similarity measure was selected to derive the clusters.

3.2.5 ***High resolution splice site analysis of first intron and splice site mutants***

3.2.5.1 *RNA isolation and cDNA synthesis*

Grain from putative splice site mutants, *int-a*.59 and *int-a*.88 and wild-type cultivars Bonus and Kristina were germinated in compost plugs at room temperature. 50—100 mg fresh leaf tissue was harvested after 7 days growth and snap frozen in liquid nitrogen. Prior to RNA extraction, tissue was ground to a fine powder using a micropestle. Total RNA extraction used the RNeasy plant mini kit (QIAGEN, cat: 74903) as per the manufacturer's protocol (http://www.qiagen.com/gb/resources/resourcedetail?id=14e7cf6e-521a-4cf7-8cbc-bf9f6fa33e24&lang=en). 5 µg of extracted RNA was reverse transcribed to cDNA using

Ready-To-Go You-Prime First-Strand Beads (Roche). RNA was incubated at 65^{o}C for 10 minutes, chilled on ice for 2 minutes; 2 µM oligoDT and reverse transcription beads were then mixed with the RNA. The reverse transcription reaction proceeded with incubation at room-temperature for 1 minute, followed by incubation at 37^{o}C for 1hr. cDNA was diluted to a final volume of 100 µl.

3.2.5.2 *Differential splice variant primer design*

Primers were designed to span the splice site in which mutations had been identified, intron 3-*int-a*.88 and intron 7-*int-a*.59. In addition, primers were designed to span intron 1, to confirm the presence of the predicted 1.2 Kbp intron. The forward primer of each pair was fluorescently labelled with a FAM fluorescent tag at the 5' end.

Table 12: The intron spanning primers designed for splice variant analysis

Primer Name	Primer Sequence	Position
HV1F	5'FAM-CTAGACGGGCCTCCGCGACC	Spans intron 1
HV1R	5'-CCATGGCAGGATTGTCACC	
HV2F	5'FAM- GGAATTCGAGGATCCCGTTG	Spans intron 2 and 3
HV2R	5'-GGGAGACAACTAGCACTGG	
HV4F	5'FAM-CATTTAAGACATGGTATGGGATACC	Spans intron 7
HV4R	5'-CAGAAAGACAAAGTAGCTCC	

3.2.5.3 *PCR conditions and product size determination*

PCR for the amplification of the intron spanning regions used 25µl reactions comprised of : 1 µl cDNA, 0.625 U TAQ DNA polymerase (Roche), 10x PCR buffer, 4 µl dNTPs (10 mM), 0.5 µl Forward and Reverse primers (20 µM) .

PCR conditions were as follows: 94^{o}C—5minutes, 24 cycles of (94^{o}C —15 seconds, 55^{o}C—30 seconds, 72^{o}C—1 minute), 72^{o}C—7 minutes.

Prior to genotyping: 1 µl PCR product was mixed with 8.97 µl Hi-di Formamide (Thermoscientific) and 0.03 µl Genescan LIZ 500 size standard (Applied Biosystems).

To determine the size of the PCR products at high resolution, PCR products were analysed on an ABI3730 DNA analyser.

The area under the peak and hence relative proportions of PCR product were determined in GeneMapper software version 4.1 (Applied Biosystems).

3.2.6 ***Phenotyping of the vrs3 allelic series***

Alleles identified as representing an independent mutation of *vrs3* (Table 11) and parental lines were planted as part of a larger 2 replicate field trial in the spring 2013 growing season at Balruddery Farm, Dundee. Each plot comprised approximately 40 plants, spread over 2 × 2 m rows. Spike emergence, the date when 50% of tillers had 50% spike emerged from the boot, was scored thrice weekly. 5 plants per plot were harvested at Zadoks growth stage 89 (Hard-dough)[22].

Post-harvest, the numbers of tillers per plant were counted. One spike per plant was used for more detailed characterisation; the following phenotypes were recorded: number of spikelet triplets, number of lateral spikelets with grain-fill, number of awned-lateral spikelets, number of awnletted lateral spikelets, number of spikelets with awned palea, length of the spike, number of extra spikelets, and number of extra florets.

Phenotypic means were generated using General ANOVA analysis in Genstat version 14 (VSNI)[141]. The interaction between functional domain mutated, parental genetic background and the *int-a* allele was assessed using the following treatment structure: Domain/Parent/Allele. Blocking structure took into account replicates, plots and spikes in the following format: Rep/Plot/Spike_number. Residual plots were assessed for deviations from normality but none were identified. A level of 95% was used to

determine significant results. Allele phenotypic means were used for bi-plot creation in SC-Biplot 1.1.

3.2.7 ***VRS3 allelic diversity/association mapping***

3.2.7.1 *Resequencing of representative haplotypes*

Using the alignment and locus order within the Barley Genome Zipper[149] SNP 12_11107 and 12_30786 were identified as flanking 11_21217, the SNP previously found to co-segregate with the *vrs3* phenotype in the BW419*Barke F_2 population (section 2.3.4.2).

17 cultivars, representative of the allelic combinations of 12_11107, 11_21217 and 12_30786, two-rowed and six-rowed, winter and spring germplasm sub-groups were identified for full re-sequencing of *VRS3.* DNA extractions were carried out as per section 3.2.2.2. PCR and sequence alignment followed the method previously detailed in 3.2.2.4.

3.2.7.2 *KASP assay design for Vrs3 allele genotyping across wider germplasm pool*

A custom KASP assay was designed by LGC Genomics to the 281 nucleotide region surrounding the A/G SNP within intron 9 of *VRS3* (Appendix 14). Resequencing had identified this SNP as diagnostic between the *Vrs3*.x and *Vrs3*.w allele of *Vrs3*.

3.2.7.3 *Vrs3 allele genotyping of cultivated barley germplasm collection*

Individuals from the HVCC[193] and AGOUEB[184] germplasm collections (Appendix 17) were genotyped with the *Vrs3*.x/*Vrs3*.w custom KASP assay using the following reaction setup and conditions.

Genomic DNA of the cultivated germplasm collection had previously been isolated and concentration established as between 50—100 ng/µl. 8 µl genotyping reactions comprised 12—25 ng genomic DNA, 0.11 µl custom KASP assay and 4 µl 2x KASP reaction mix v.4 (LGC genomics). Reaction conditions were as follows: 20^{o}C —2 min; 94^{0}C - 15 min; 10 touch-down cycles of 94^{o}C —20 s, 62^{o}C— 1 min, -0.7^{o}C per cycle; 32 cycles of 94^{o}C —20 s, 55^{o}C— 1min; 20^{o}C —2 min. All reactions were run on a StepOnePlus real-time PCR machine (LifeTechnologies).

3.2.7.4 *Association Panel genome-wide genotyping data*

iSelect SNP genotyping information[36,184,193] was collated for all individuals genotyped with the *Vrs3*.x/*Vrs3*.w genotyping assay. The *Vrs3* allele was included as a SNP. Monomorphic SNPs, SNPs with a minor allele frequency <10% or with >20% missing values were removed from the dataset prior to GWAS.

3.2.7.5 *Principal coordinate analysis*

Principal coordinate analysis and cluster analysis were performed using PAST[192] to establish distinct populations for GWAS.

3.2.7.6 *GWAS*

GWAS used Genstat version 14[141]. In each case, two independent analyses were run, namely null, assuming no population structure and Eigenstrat, including significant principal components as co-variates. To establish the significance level at which the false discovery rate was less than 99% $-\text{Log}_{10}$ (P) values were converted to q values using the q value library script (version 1.43.0) in R[194].

3.2.8 **VRS3** ***phylogenetic tree construction***

A reciprocal BLAST of the barley VRS3 predicted protein sequence against the plant Phytozome database version 9.1[195] identified a family of 53 protein sequences from the Embryophyte family with a bit score of 824.3 and an E value of 0.0. Domain identification of the identified 53 protein sequences found 5 to have an additional FY/R rich C Terminal domain, these were therefore discounted from further analysis. Additionally sequences which were not predicted to contain all three functional domains identified within VRS3 were removed prior to sequence alignment.

A further BLAST of the wheat survey sequence with barley genomic sequence of *Vrs3* identified 3 orthologues on chromosomes 1A, 1B and 1D[196]. Predicted protein sequences of the wheat VRS3 orthologues were established by alignment with the *Vrs3* barley FlcDNA (AK368672) to establish reading frame and then translation into peptide sequence.

BLAST of VRS3 against Morex version 3 predicted protein sequences identified an orthologue of VRS3, MLOC_53868.1 on chromosome 6H (Bit score: 709, E value: 0.0). BLAST of this sequence against the wheat survey sequence identified homologous sequence on chromosome 1A and 1D, however, no reliable sequence could be

established for chromosome 1B. Alignment with MLOC_53868.1 established reading frame for the orthologous wheat sequence prior to translation.

All protein sequences orthologous to VRS3 were aligned using MUSCLE[188] alignment algorithm.

Phylogenetic tree was constructed using a Maximum Likelihood model in MEGA version 6[197] with 100 bootstraps.

3.2.9 ***Amino acid conservation score***

Orthologous VRS3 sequences identified in 3.2.8 were aligned using MUSCLE[188]. Protein residue conservation prediction software[198] was used to derive a Shannon Entropy conservation score for each VRS3 amino acid residue.

3.3 *Results*

3.3.1 *Prioritisation of the investigation of candidate genes for the identification of VRS3.*

A review of previously published gene family functions and phenotypes was undertaken to establish an order of priority for the investigation of the 16 identified candidates genes within the *VRS3* mapping interval (Table 13).

Table 13: Putative functions and associated mutant phenotypes of the sixteen candidate genes identified in the mapping of *VRS3.*
The locus marked in bold has been identified as the primary candidate.

Barley SNP	Rice Locus and description	Putative Gene Function	Associated Mutant phenotype	Ref.
11_10520	LOC_Os10g42610 expressed protein			
	LOC_Os10g42620 dihydroflavonol-4-reductase	Production of anthocyanin via Flavonoid biosynthesis pathway	Loss of pigmentation	199
	LOC_Os10g42630 expressed protein	unknown	unknown	
	LOC_Os10g42660 expressed protein	unknown	unknown	
	LOC_Os10g42670 glycosyl hydrolases family 16	Glycosidic bond formation in cell walls (Barley)	Changes in grain beta-glucan content	200,201
	LOC_Os10g42690 jmjC domain containing protein	**OsJMJ706 Histone demethylation (Rice)**	**Altered floral morphology and organ number**	202
	LOC_Os10g42700 CS domain containing protein	Interaction with heat shock protein 90 (HSP90)	Breakdown of pathogen resistance	203 204
11_21217	LOC_Os10g42710 *RCD1* (Radical Induced Cell Death 1)	Regulation of cell death (Arabidopsis)	Programmed cell death on exposure to reactive oxygen species	205,206
	LOC_Os10g42720 acyltransferase	Catalysis of glycerol 3-phosphate to 1-acyl-glycerol-3-phosphate	Alteration in chilling tolerance	207,208
	LOC_Os10g42724 VHS and GAT domain containing protein	Intracellular vesicle trafficking (Rice)	unknown	209

	LOC_Os10g42750 CSLD1 - cellulose synthase-like family D	Root hair elongation (Rice)	Short root hairs	210
	LOC_Os10g42780 lrgB-like family protein	Chloroplast inner membrane protein (Arabidopsis) exact function unknown	Chlorotic regions on leaves	211,212
	LOC_Os10g42790 cytidylyltransferase domain containing protein	Phosphatidylcholine biosynthesis pathway, cold tolerance (Arabidopsis)	No phenotype at ambient temperature	213 214
	LOC_Os10g42800 AMP-binding enzyme	4-coumarate-CoA ligase, Flavanoid and lignin biosynthesis pathways	Reduced lignin content	215,216
	LOC_Os10g42820 early-responsive to dehydration protein-related	Localised to chloroplast membrane (Arabidopsis) function unknown	unknown	217
	LOC_Os10g42830 transporter family protein	Orthologous to vacuolar glucose transporter 1 (Arabidopsis)	Delayed flowering and reduced germination in Arabidopsis mutants	218
	LOC_Os10g42840 NADH dehydrogenase 1 alpha subcomplex subunit 12	Proton pump across mitochondrial membrane (Maize, Tobacco)	Male sterility in maize and tobacco	219,220
12_30786	LOC_Os10g42860 ribosomal protein L35			

Of the sixteen loci, LOC_Os10g42690, a JmjC histone demethylase of rice, has previously been shown to result in altered floral morphology, in particular loss or gain of lemma and palea, when mutated in rice.[202] LOC_Os10g42690 was therefore deemed the priority candidate for further investigation into the gene underlying the *Vrs3* locus.

3.3.2 ***A JmjC domain containing candidate for Vrs3***

3.3.2.1 *Predicted gene structure*

Alignment of genomic DNA contigs: 5669 (Morex) and 845567 (Bowman) with FlcDNA contig (AK368672) predicted the coding sequence of the JmjC gene candidate to comprise 4868 nucleotides, structured over 11 exons with a 1.2 kbp intron between the predicted translational start codon and next amino-acid nucleotide triplet.

3.3.2.2 *JmjC candidate resequencing*

The JmjC candidate was resequenced in BW902 and BW419, the Bowman isolines containing *vrs3* mutant introgressions *vrs3*.f and *int-a*.1 respectively; Bowman, the isoline recurrent parent; Hakata 2 and Bonus , the parent cultivars of the original mutant's *vrs3*.f and *int-a*.1 respectively and Morex for comparison to the genomic reference sequence. Sequencing of BW419 identified a 2 bp deletion within exon 9, 3468bp from the start codon. Re-sequencing the region of exon 9 in 42 randomly selected F_2 individuals from the BW419Bo population (Chapter 2) found the presence/absence of the deletion to correlate 100% with the mutant/two-rowed phenotype, providing strong evidence that this polymorphism underlies the observed *vrs3* phenotype in BW419.

Resequencing of the JmjC candidate proceeded across the 32 accessions from the *vrs3* allelic series and their parental cultivars Foma, Kristina and Hege. A single base pair deletion was identified in *vrs3*.f and BW902 but absent in Bowman and Hakata 2, consistent with this mutation underlying the phenotype observed in the *vrs3*.f and BW902 mutants. Polymorphisms were identified in 31 accessions from the *vrs3* allelic series when compared to sequence of their parental cultivars. The exception being *int-a*.37 and it may be that mutation in this accession lies 5' in upstream regulatory elements of the region of *vrs3* sequenced. The identification of 31 mutational events within the JmjC candidate in these mutant lines provided extremely strong evidence that the gene underlying the *VRS3* locus is an orthologue of the rice JMJD2 histone demethylase *JMJ706* (LOC_Os10g42690).

Table 14 details the positions and types of polymorphisms identified in each of the alleles. In total, 3 frameshift, 17 missense, 2 splice-site and 9 nonsense mutations were identified. It is interesting to note that all frameshift mutations were the result of irradiation by either X or Gamma-rays (Table 11) which is consistent with previous findings that physical mutagenesis by radiation result in DNA break/repair and chemical mutagenesis nucleotide point mutation through mechanisms such as alkylation[221].

Table 14: A summary of the identified mutations within the *int-a* allelic series

Allele	Background Cultivar	Mutagen	Mutation	Mutation Type	Reference Amino Acid
int-a.27	Foma	n-PMS	G1769A	missense-SNP	G131D
int-a.30	Foma	n-PMS	G1769T	missense-SNP	G131V
int-a.54	Kristina	EMS	G1769A	missense-SNP	G131D
int-a.55	Kristina	IPMS	G1769A	missense-SNP	G131D
int-a.71	Bonus	SA	G1769A	missense-SNP	G131D
BW902	Bowman	Gamma-ray	T1774-	Frameshift	C133-
vrs3.f	Hakata 2	Gamma-ray	T1774-	Frameshift	C133-
int-a.64	Bonus	IPMS	CT1856AA	missense-SNP	T160K
int-a.88	Bonus	SA	G1926A	Splice site mutation	
int-a.46	Kristina	Neutrons	G2139T	missense-SNP	V229F
int-a.79	Bonus	SA	G2448A	missense-SNP	G275E
int-a.31	Foma	Neutrons Neutrons EMS Neutrons	C2465T	missense-SNP	L281F
int-a.32	Foma	EMS	C2465T	missense-SNP	L281F
int-a.61	Bonus*	EMS	C2465T	missense-SNP	L281F
int-a.51	Kristina	IPMS	T2468A	missense-SNP	Y282N
int-a.2	Bonus	Radon	G2474C	missense-SNP	G284R
int-a.34	Foma	EI	C2524A	nonsense-SNP	Y300STOP
int-a.35	Foma	PDSADEE	C2524A	nonsense-SNP	Y300STOP
int-a.9	Foma*	EI	A2525T	missense-SNP	S301W
int-a.14	Foma	DES	A2525T	missense-SNP	S301W
int-a.52	Kristina	X-rays	A2629G	missense-SNP	H306R
int-a.77	Bonus	EMS	G2679T	nonsense-SNP	G323STOP
int-a.74	Bonus	EMS	A2688T	nonsense-SNP	K326STOP
int-a.17	Foma	EI	A2823T	nonsense-SNP	K371STOP
int-a.59	unknown*	SA	G2893A	Splice site mutation	
int-a.10	Bonus	EI	A2993G	missense-SNP	E399G
int-a.102	Hege	SA	G2995A	missense-SNP	A400T
int-a.103	Hege	SA	G2995A	missense-SNP	A400T
int-a.21	Foma	EMS	G3021A	nonsense-SNP	W408STOP
int-a.8	Bonus	X-rays	AT3288--	Frameshift	LS497--
BW419	Bowman	X-rays	CG3464--	Frameshift	R529-
int-a.1	Bonus	X-rays	CG3464--	Frameshift	R529-
int-a.86	Bonus	SA	G3471A	missense-SNP	C531Y
int-a.37	Foma	Neutrons	None identified	None identified	-

*denotes those alleles found to be in a different genetic background to previously published. EI=Ethylene Imine, DES= diethyl sulfate, EMS= ethyl methanesulfonate, n-PMS = n-propyl methanesulfonate, PDSADEE=propanedisulfonic acid diethylester, IPMS= iso-propyl methanesulfonate, SA= Sodium azide

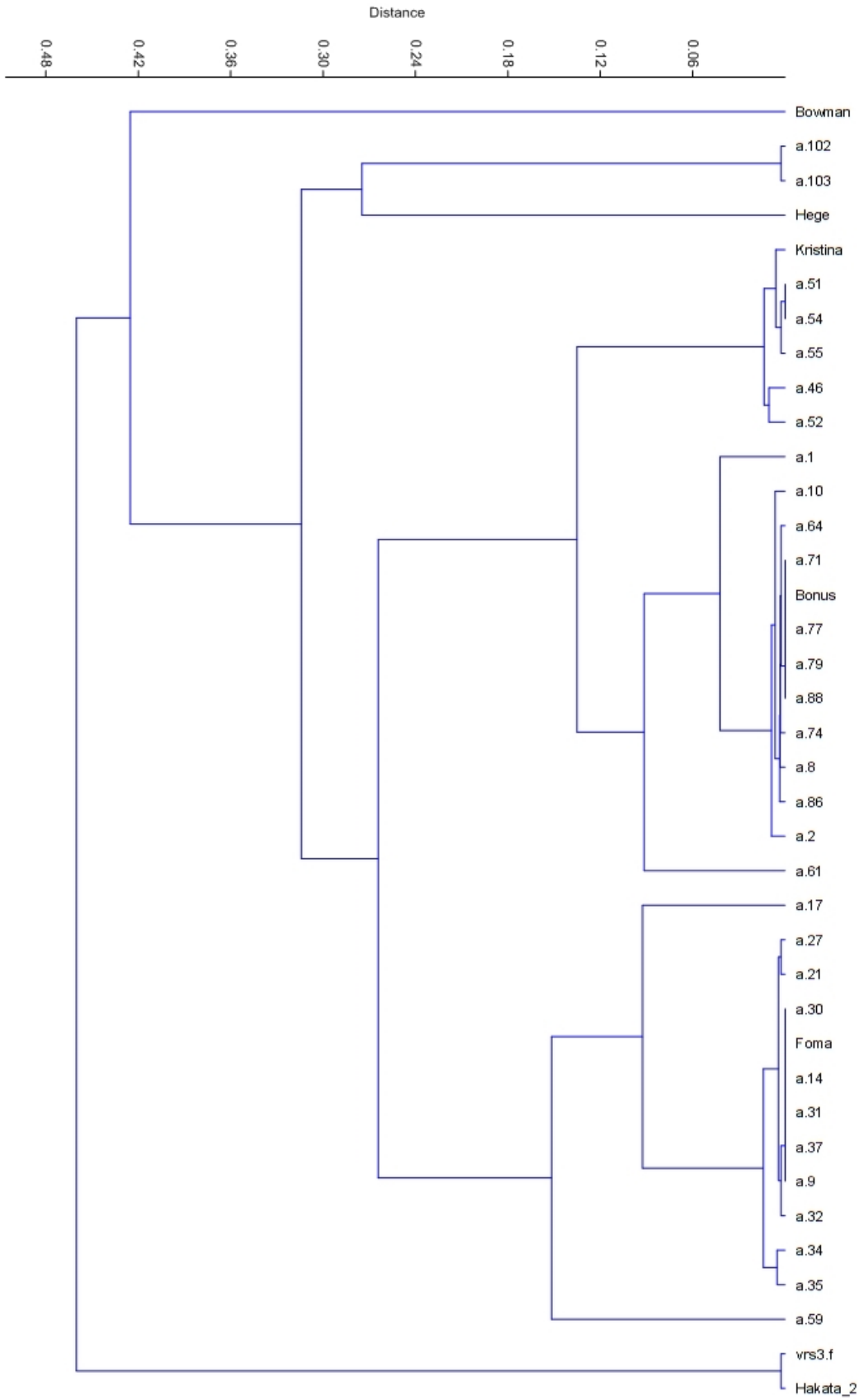


Figure 24: Dendrogram illustrating the genetic distance between the *vrs3* allelic series and parent cultivars.

The 31 identified mutations covered only 17 unique positions within *Vrs3*. For example the nucleotide 1769 base pairs from the start codon is mutagenised in 5 *vrs3* accessions (Table 14). Of the 5 accessions, 4 share the same G/A polymorphism, with only *int-a*.30 encoding an alternative G/T polymorphism. To help confirm these as independent mutational events, individual mutant alleles were genotyped with the Illumina 384-BeadXpress XOPA[142]. Whilst we would expect no differences between mutants from the same parent cultivars, Figure 24 shows clear separation between alleles and parents in some cases.

Three of the four G/A polymorphisms at nucleotide 1769 are within different genetic backgrounds, negating the possibility of multiple selections from the same original mutational event. This would suggest that either some regions of the gene are more susceptible to mutagenesis than others or that these nucleotides are key to the function of the gene and therefore their mutagenesis results in an observable, non-lethal phenotype in the plant. However, in some cases the results do suggest duplicate selection of the same mutational event with the following pairs of mutants sharing identical mutations in the same genetic backgrounds: *int-a*.9 and *int-a*.14, *int-a*.31 and *int-a*.32, *int-a*.34 and int-35, *int-a*.54 and *int-a*.55, and *int-a*.102 and *int-a*.103.

In addition to the causal mutations within the induced *vrs3* allelic series, sequencing of *Vrs3* identified a further six SNPs which separated the parental cultivars into two groups with Bonus and Hakata 2 forming one group and the remaining four cultivars a second group (data not shown). The occurrence of these natural alleles of *Vrs3* will be discussed in more detail later. However, the segregation of the natural alleles in combination with the genetic relationships illustrated in Figure 24 suggested that *int-a*.59 is not induced in the genetic background Kristina as previously documented. *int-a*.9 and *int-a*.61 also showed discrepancies with regard to the genotype in which they are documented to have been induced, with genome-wide genotyping suggesting *int-a*.9 to be induced in Foma rather than Bonus and *int-a*.61 having a closer genetic relationship to Bonus than Kristina. Additionally, Figure 24 suggests whilst *int-a*.102 and *int-a*.103 are closely related, Hege, the parent in which they are both induced is more distantly related when compared to the relationship of other parent cultivars and their respective induced alleles. This indicates that the parent Hege was probably not genetically uniform and that a difference subline was used for mutation compared to the one we used for genotyping.

3.3.3 *Vrs3 gene structure*

Interpro[222] analysis of the barley VRS3 predicted protein sequence, translated from the flcDNA, identified three functional domains, JmjN, JmjC and C5HC2 Zinc finger consistent with the previous findings in rice[202]. The JmjN and JmjC domains of VRS3 share homology across species; BLASTp searches of the NCBI protein database[223] with the 33 amino acids of the VRS3 JmjN domain and 124 amino acids of the JmjC domain found the top 100 hits to have 80% and 86% identity scores respectively. The VRS3 C5HC2 Zinc finger (ZnF) domain is less well conserved; the top 100 BLASTp hits averaged 54% identity.

Phylogenetic studies of the rice JMJ706 suggest that it falls within the JMJD2 clade of histone demethylases[202]. This class of histone demethylases is characterised by the presence of the JmjN domain followed directly by the JmjC domain[148]. Barley VRS3 shares 73% overall protein sequence identity with rice JMJ706. This increases to 88%, 96% and 80% respectively when only comparison of the JmjN, JmjC and C5HC2 ZnF functional domains are considered, suggesting that *Vrs3* too is likely to be part of the JMJD2 clade.

The predicted JmJN domain of *Vrs3* spans 99 nucleotides (33 amino acids), 1596 nucleotides 3' from the start codon. Studies in yeast and humans have suggested the JmjN domain stabilises the protein structure through interaction with the JmjC domain via beta-strands[224]. Alignment of the JmjN and JmjC domains from rice and barley with human JMJD2D protein sequence found that the *vrs3*.f deletion mutant is within a conserved beta strand motif (Figure **25**).

The predicted JmjC domain of *Vrs3* spans 372 nucleotides (124 amino acids), 2462 nucleotides 3' of the start codon. The structure of the JmjC domain of JMJD2 histone demethylases has been previously studied in detail[225]. The demethylase reaction catalysed by JmjC histone demethylases requires Fe^{2+} and alpha-ketoglutarate as cofactors[226]. These co-factors are bound by highly conserved residues within the demethylase domain[225].

Figure 25 shows the position of these conserved residues within the JmjC domain. None of the 5 cofactor binding residues are mutated within the *vrs3* mutant allelic series.

The predicted C5HC2 ZnF domain of *Vrs3* spans 162 nucleotides (54 amino acids), 3292 nucleotides 3' of the start codon. Previous studies of the rice JMJ706 protein,

found the deletion of the C5HC2 domain resulted in a change from nuclear to cytoplasmic localisation of the protein in transfected onion cells, suggesting a role of C5HC2 in cellular localisation[202].

3.3.3.1 *vrs3 allelic series mutations relative to functional domains*

Figure 26 illustrates the predicted structure of *VRS3* and the positions of the identified *vrs3* allelic series polymorphisms relative to the predicted functional domains. It is apparent that the majority of the mutant alleles are found either within or in close proximity to the three functional domains of *VRS3*. This would suggest that the nucleotides substituted are key to *VRS3* functionality without being lethal to the plant.

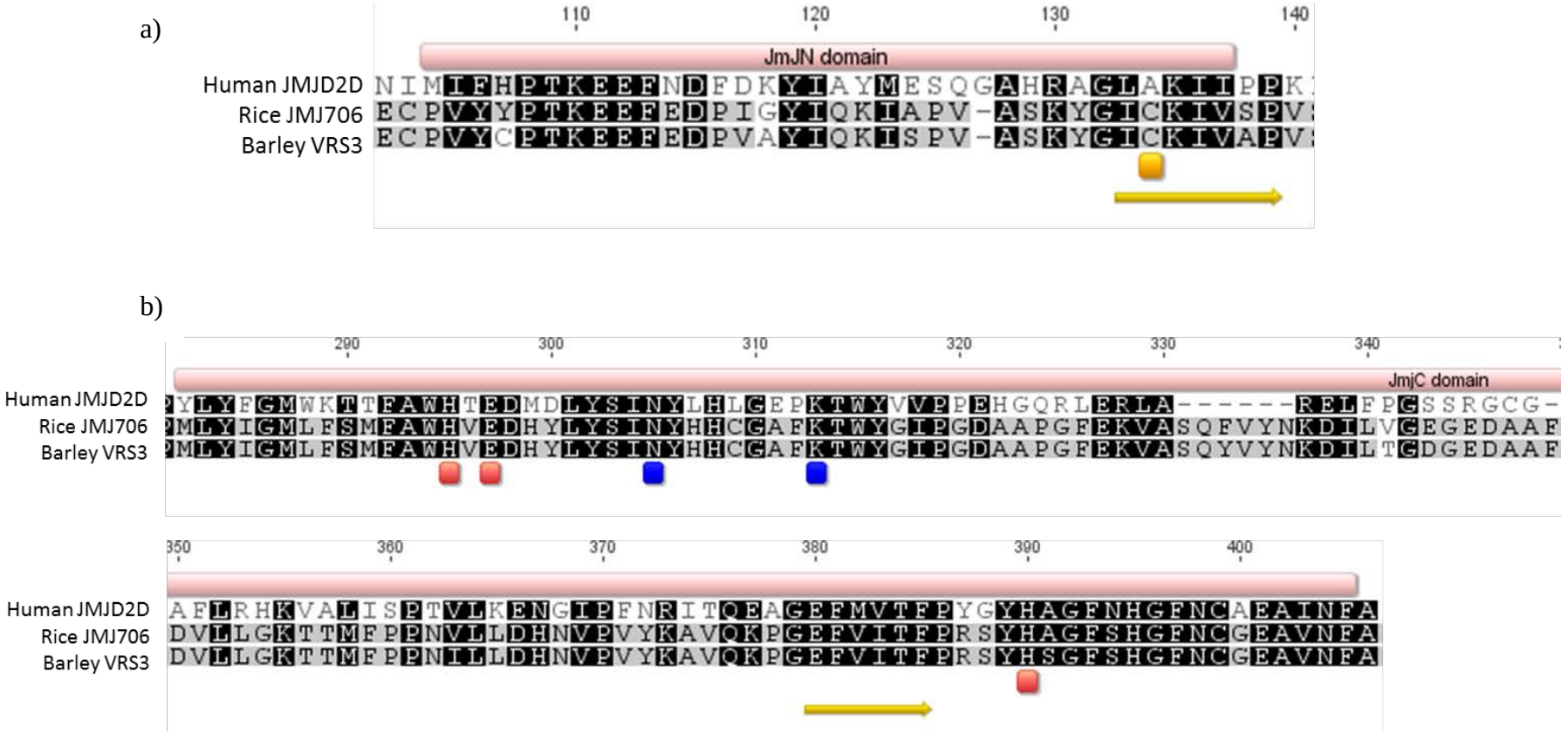


Figure 25 Alignment of the conserved a) JmjN and b) JmjC domains from human (JMJD2D), rice (JMJ706) and barley (VRS3).

Yellow arrows highlight the conserved beta-strands, with the yellow square indicating the position of the vrs3.f mutation. The red squares and blue squares indicate conserved co-factor binding residues, red-alpha ketoglutarate binding and blue-Fe^{2+} binding.

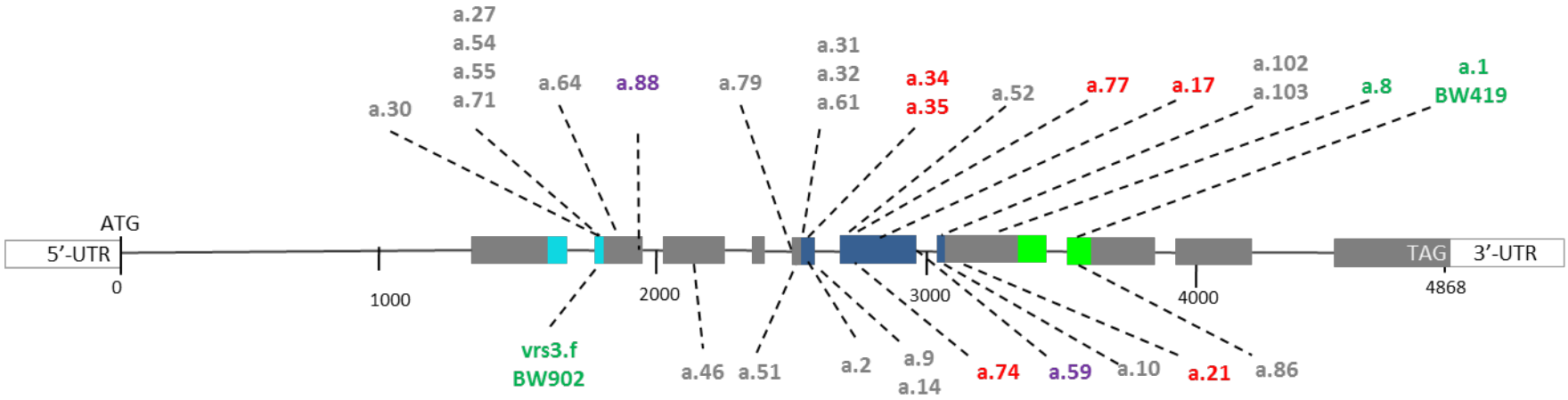


Figure 26: *VRS3* gene structure illustrating the position of the predicted functional domains and their positions relative to the 11 exons and identified *vrs3* allelic series.

Exons are represented by dark grey rectangles. The turquoise region spanning exon2 and 3 represents the predicted JmjN domain, the dark blue region spanning exons 6, 7 and 8 represents the predicted JmjC domain and the green region spanning exons 8 and 9 represents the predicted Zinc Finger domain. Frameshift mutants are illustrated in green, nonsense mutants are illustrated in red, missense mutants are illustrated in grey, splice site mutants are shown in purple.

Frame-shift and nonsense mutations

Translation of the nucleotide sequence of the frame-shift and nonsense mutation containing alleles predicts them to form a deletion series across the VRS3 protein (Figure 27). The *vrs3*.f allele is predicted to result in the largest deletion with all three domains, JmjN, JmjC and ZnF affected. In contrast the *int-a*.1 allele results in deletion of a portion of the ZnF domain and the C-terminal region. Despite the *vrs3*.f allele likely to result in the complete knock-out of the VRS3 protein, the mutation is not lethal to the plant suggesting that VRS3 function is not vital for plant survival.

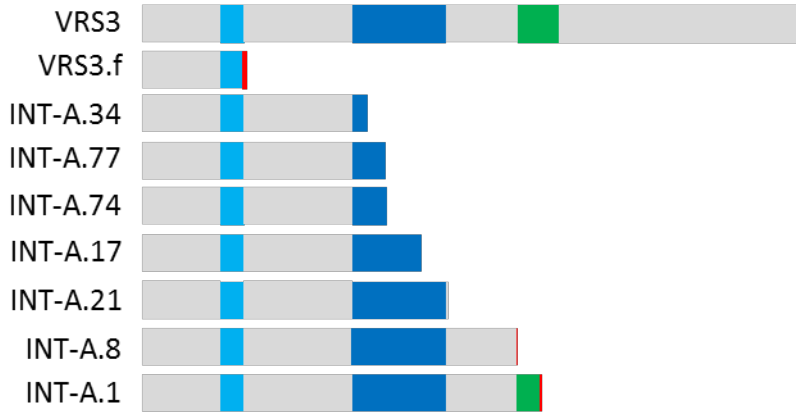


Figure 27 Cartoon representation of the wild-type protein VRS3 and the deletion series predicted from the frame-shift and nonsense mutation *vrs3* alleles.
The three functional domains are illustrated JmjN(turquoise), JmjC(dark blue) and ZnF(green). Regions of non-synonymous amino-acids due to frameshift are illustrated in red.

Missense mutations

Non-synonymous substitutions within the *vrs3* allelic series result in 13 missense mutations, largely within the predicted functional domains of *Vrs3*. In all cases the substituted amino acid has different chemical properties compared to wild-type (Table 15) suggesting disruption to conformation of the VRS3 protein is highly likely. For example the substitution of a low molecular weight amino acid for a high weight amino such as glycine to asparagine in the case of *int-a*.27 is likely to result in conformational change to the protein due to the size of the amino acid. Cysteine residues are key components of the zinc finger domain, therefore the substitution in the *int- a*.86 allele of a cysteine residue capable of forming strong permanent covalent disulphide bonds for a tyrosine residue which is characterised by a hydroxyl residue capable of only weaker hydrogen bonding is likely to result in altered functionality of the VRS3 protein.

Table 15 : Summary of the differences in chemical character between wild-type and substituted amino acid in the missense mutations of the vrs3 allelic series.

Allele	Wild-type amino acid	Missense mutant amino acid
a.27 a.54 a.71	**Glycine** Neutrally charged Low Molecular Weight Neutral hydrophobicity	**Asparagine** Polar Higher molecular weight Hydrophilic
a.30	**Glycine** Neutral hydrophobicity	**Valine** Hydrophobic
a.64	**Threonine** Polar Neutral hydrophobicity	**Lysine** Positively charged Hydrophilic
a.46	**Valine** Aliphatic Low molecular weight	**Phenylalanine** Aromatic High molecular weight
a.79	**Glycine** Neutrally charged Low Molecular Weight Neutral hydrophobicity	**Glutamic Acid** Negatively charged Higher molecular weight Hydrophilic
a.31 a.61	**Leucine** Aliphatic	**Phenylalanine** Aromatic
a.51	**Tyrosine** Aromatic Hydroxyl	**Asparagine** Aliphatic Amide
a.2	**Glycine** Neutrally charged Low Molecular Weight Neutral hydrophobicity	**Arginine** Positively Charged High Molecular Weight Hydrophilic
a.9 a.14	**Serine** Polar Aliphatic Neutral hydrophobicity Low molecular weight	**Tryptophan** Neutral Charge Aliphatic Hydrophobic High Molecular weight
a.52	**Histidine** Aromatic	**Arginine** Aliphatic
a.10	**Glutamic Acid** Negative charge High molecular weight	**Glycine** Neutrally charged Low Molecular Weight
a.102	**Alanine** Neutral Charge Hydrophobic	**Threonine** Polar Neutral hydrophobicity
a.86	**Cysteine** Disulphide bond formation Aliphatic	**Tyrosine** Hydroxyl group Aromatic

3.3.4 *vrs3 splice site mutants*

Of the mutations identified within the *vrs3* mutant allelic series, two, *int-a*.88 and *int-a*.59 are within the intron rather than exon region of *vrs3*. The relative position of their respective mutations suggested they could be potential splice site mutations.

For intron recognition and successful splicing, the spliceosome complex relies on four conserved motifs within the intron, a GU nucleotide motif at the 5' intron boundary, an AU nucleotide motif and polypyrimidine tract at the 3' intron boundary and an A residue within the intron upstream from the 3' intron boundary[227]. If conserved splicing motifs are disrupted, the spliceosome complex seeks to either find alternative splice sites in the region or alternatively no splicing occurs and the intron is retained[228]. Sequencing of *vrs3* in *int-a*.88 and *int-a*.59 identified polymorphism which appeared to disrupt the conserved GU motif required for splicing (Figure 28).

High resolution splice site analysis[229] was used to further establish *int-a*.88 and *int-a*.59 as splice site mutants (Figure 29). This technique uses fluorescently-labelled, intron-spanning primers to amplify across the splice site of interest from cDNA. Following amplification and fragment separation using a DNA fragment analyser, the relative sizes and proportions of amplified fragments are determined. Comparison of amplicon sizes across the intron enables the prediction of splicing events.

Figure 28: Comparison of the nucleotide sequence between Bonus and Kristina, and putative *int-a* splice site mutants, *int-a.88* and *int-a.59*.
Exons are identified by black shading, with the causal polymorphism highlighted in orange.

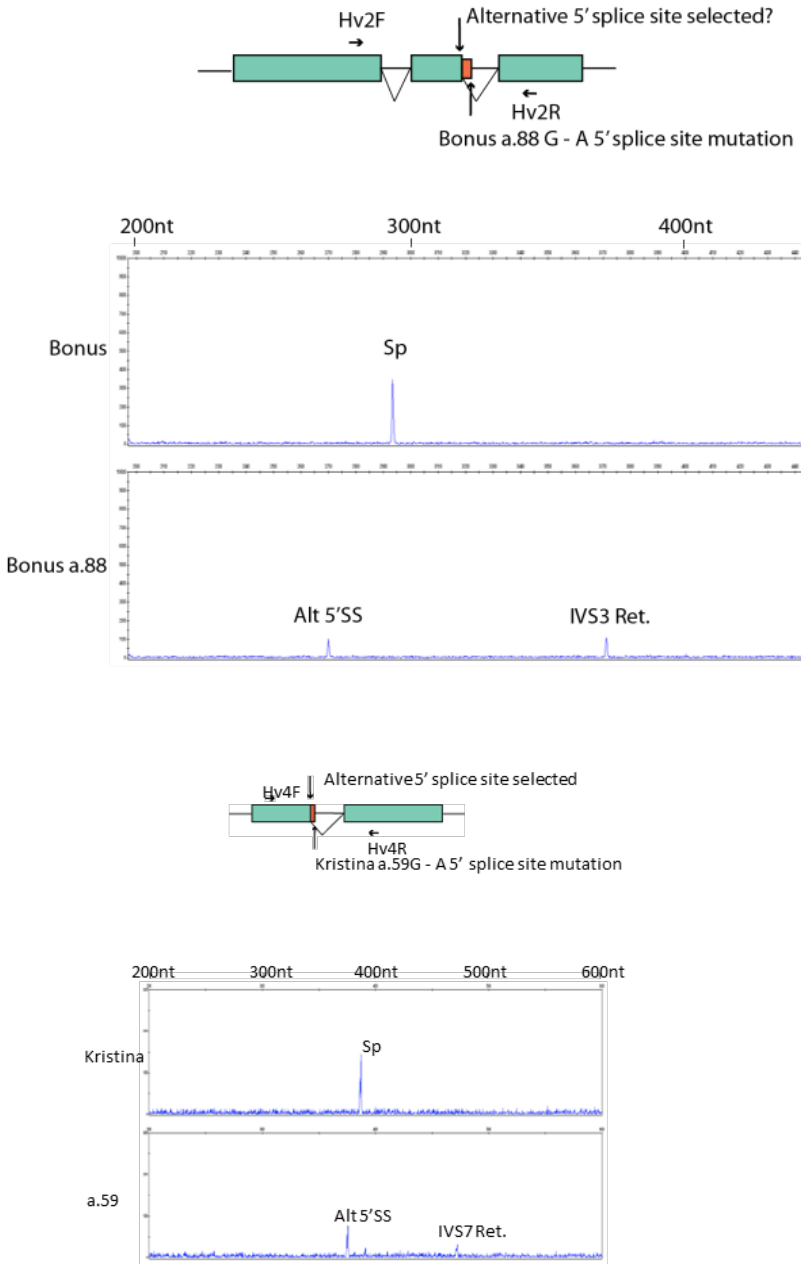


Figure 29:High resolution analysis of the splice site mutants int-a.88 and int-a.59.
The gene structures illustrated in green show the relative positions of the intron spanning primers and affected splice sites. The graphs below the gene structures show the size peak of the amplified PCR product in wild-type (Bonus) and respective mutants. Sp marks the wild-type spliced product; Alt SS the alternative spliced product and IVS the product with intron retained.

Wild-type splicing of *Vrs3* across intron 3, resulted in a single large peak of 294 bp consistent with the expected splicing event. In the *int-a*.88 mutant no peak was observed at 294 bp but two smaller peaks observed at 270 bp and 371 bp, which would suggest the use of an alternative 5' upstream splice site and the retention of intron 3 respectively. The peaks were of similar area suggesting, both scenarios to occur with similar frequency. Prediction of the amino-acid sequence of the two mutant splice products suggests that both intron retention and use of the alternative 5'splice site in the *int-a*.88 mutant would result in the truncation of the VRS3 peptide sequence due to the introduction of a premature stop codon in both cases (Appendix 16).

Analysis of the splicing across intron 7 found wild-type splicing resulted in a single peak of 387 bp; the splice site mutant *int-a*.59 was found to produce two different sized spliced products 375 bp and 472 bp, which suggests both use of an alternative 5' splice site and retention of intron 7 in this mutant. The area of the peak at 472 bp was 1/3 the area of the peak at 375 bp suggesting predominant use of an alternative 5' splice site in the *int-a*.59 mutant. Both of these splicing events would result in disruption of the C terminal region of the JmjC domain. In this case the retention of intron 7 is predicted to result in a truncated protein due to the introduction of a premature STOP codon at the C-terminal of the JmjC domain. The use of the alternative 5' splice site in the *int-a*.59 allele is predicted to maintain the frame of the protein but result in the deletion of four amino acids at the C-terminal of the JmjC domain (Appendix 16). Further analysis would be required to ascertain what effect this deletion may have on the functionality of the VRS3 protein.

The presence of a large intron (approximately 1.2 kb) between the predicted translation start codon and next amino acid coding codon is an unusual feature of the *Vrs3* gene structure. To confirm its presence, intron spanning primers were used to amplify splice products from this region using total-RNA which had previously been isolated from both barley leaf and inflorescence tissue. High resolution analysis of these products is illustrated in Figure 30. The predominant splice product is 165 bp, which is the expected size for the presence of a 1.2 kb intron. There is evidence of shorter and longer spliced products in both tissues but these are at very low levels compared to the main spliced product. More detailed cloning combined with sequencing analysis would be needed to confirm the exact nature of these additional spliced products.

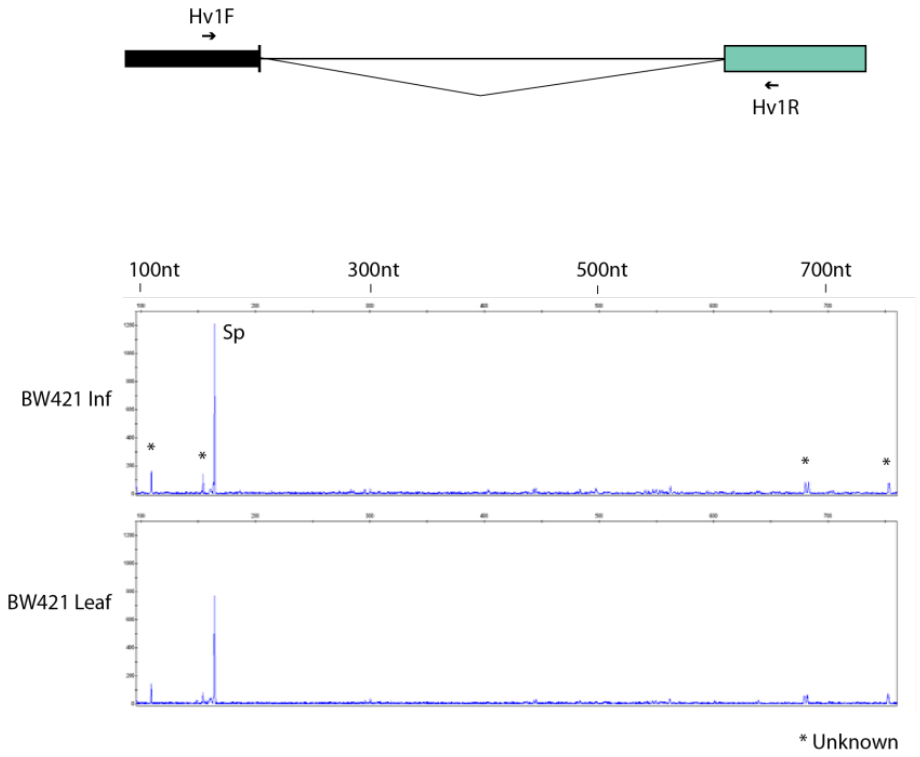


Figure 30: High resolution analysis of intron 1 splicing

*The start codon is illustrated as a vertical black bar in the gene structure and the second exon is, illustrated in green. The primary spliced product is identified as Sp and the unknown products are designated by *.*

3.3.5 *vrs3 induced mutant allelic series relationships between phenotypes and genotypes*

The predominant features of the *vrs3* mutant phenotype have been previously described in (section 2.3.1). Figure 31 shows a representative spike of each *vrs3* mutant allele, compared to its two-rowed parent. It was presumed that those allelic series accessions which had identical mutations in the same genetic background, e.g. *int-a*.34 and *int-a*.35, represented multiple selections of the same original mutation and therefore in these cases only a single representative was used. It is evident that the strength of phenotype varied greatly across the differing *vrs3* accessions, with some accessions e.g. *int-a*.17 and *int-a*.31 having a more two-rowed phenotype and *int-a*.54 and *vrs3*.f, a more six-rowed phenotype.

Figure 31: Representative spikes of the *vrs3* allelic series

30i. two-rowed cultivar Bonus a) in comparison with Bonus induced mutants; b) int-a.1, c)int-a.2, d) int-a.8, e) int-a.10, f) int-a.61, g) int-a.64, h) int-a.71, i) int-a.74, j) int-a.77, k) int-a.86, l) int- a.88

Figure 31*i Continued*

Figure 31ii: *Comparison of a) two-rowed cultivar Kristina with b) int-a.46, c) int-a.52, and d) int-a.54*

Figure 31iii*: Comparison of a) two-rowed cultivar Foma with b) int-a.9, c) int-a.17, d) int-a.21, e) int-a.27, f) int-a.30, g)int-a.31, h) int-a.34*

Figure 31*iv: Comparison of a) two-rowed cultivar Hakata2 with b) vrs3.f and c) two-rowed cultivar Hege with d) int-a.102, e) int- a.59 of unknown parentage.*

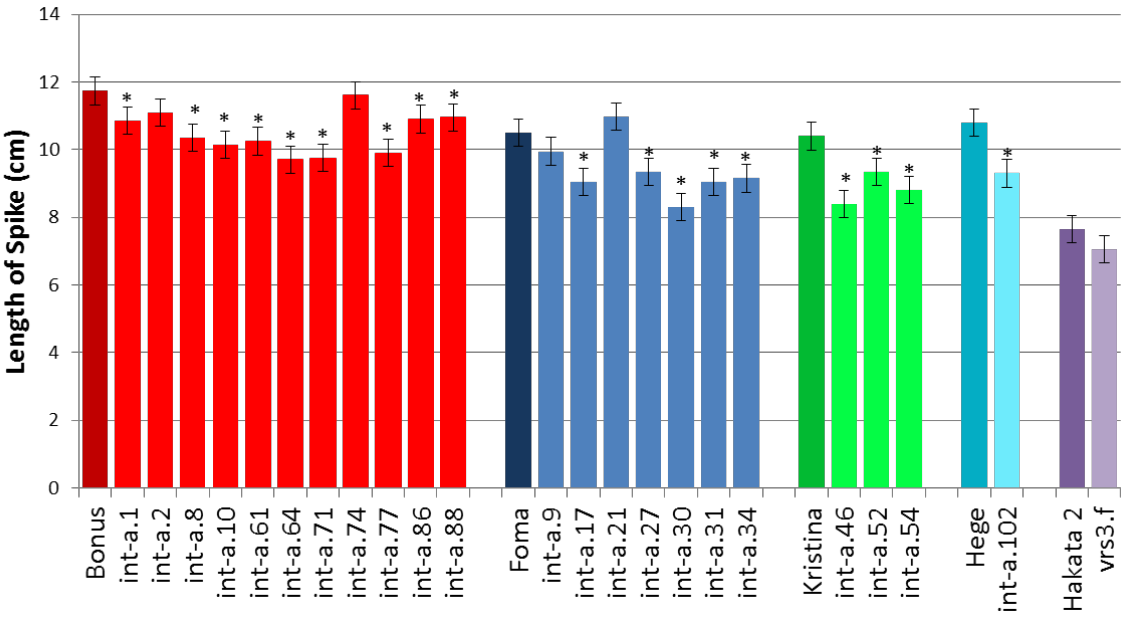
Figure 31v: *Comparison of a) two-rowed Bowman with b) BW902 (vrs3.f*Bowman (BC6)) and c) BW419 (int-a.1*Bowman (BC6))*

3.3.5.1 *Quantitative phenotypic diversity across the vrs3 allelic series*

General ANOVA analysis was used to establish significant phenotypic differences between *vrs3* alleles and wild-type parents. Analysis between *VRS3* functional domains was also used to try and establish any associations between mutant phenotypes and specific regions of *VRS3.*

Spike length and spikelet density

a)



b)

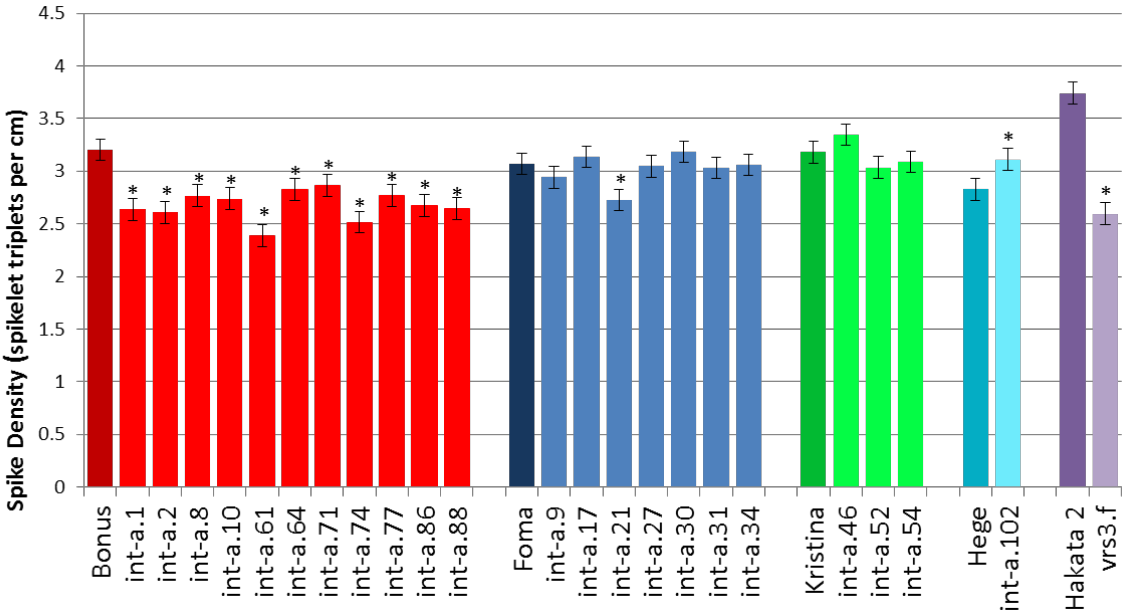


Figure 32:Mean a) Spike length and b) spike density across the vrs3 allelic series
*Colours represent alleles in the same genetic background with the parental genotype highlighted in a darker shade. Error bars show +/- SED. * represents those alleles with mean significantly different from their respective parental genotype.*

Spike length was affected by mutation at *vrs3 (*p=0.001), with the majority of alleles producing significantly shorter spikes than their respective wild-type parent (Figure 32). Comparison across domains, found the shortest mean spike length, 8.65 cm, to be

associated with mutations of the JmJN domain; this was significantly shorter than the means of the other domains (p=<0.001).

In addition to changes in spike length, *vrs3* alleles were also associated with a difference in spikelet density (spikelet triplets per cm of spike) (p=0.026) (Figure 32). This difference was most pronounced in alleles from the Bonus background; all Bonus *vrs3* alleles were found to have significantly lower spikelet density than Bonus, even in cases, such as *int-a*.2, where no significant difference in overall spike length had been found. However, this was not the case for the alleles originating from a Foma or Kristina background, as in these cases the majority of alleles showed no significant difference from their respective wild-type parents; the exceptions to this were *int-a*.21 which showed significantly lower spikelet density and *int-a*.46 which was the only allele to show a significantly greater spikelet density compared to wild-type. Comparison across domains found mutations of all domains to produce a mean spikelet density lower than wild-type, with mutations of the ZnF domain producing significantly laxer spikes compared to mutations of the JmjC and JmjN domains (p<0.001).

Spike emergence

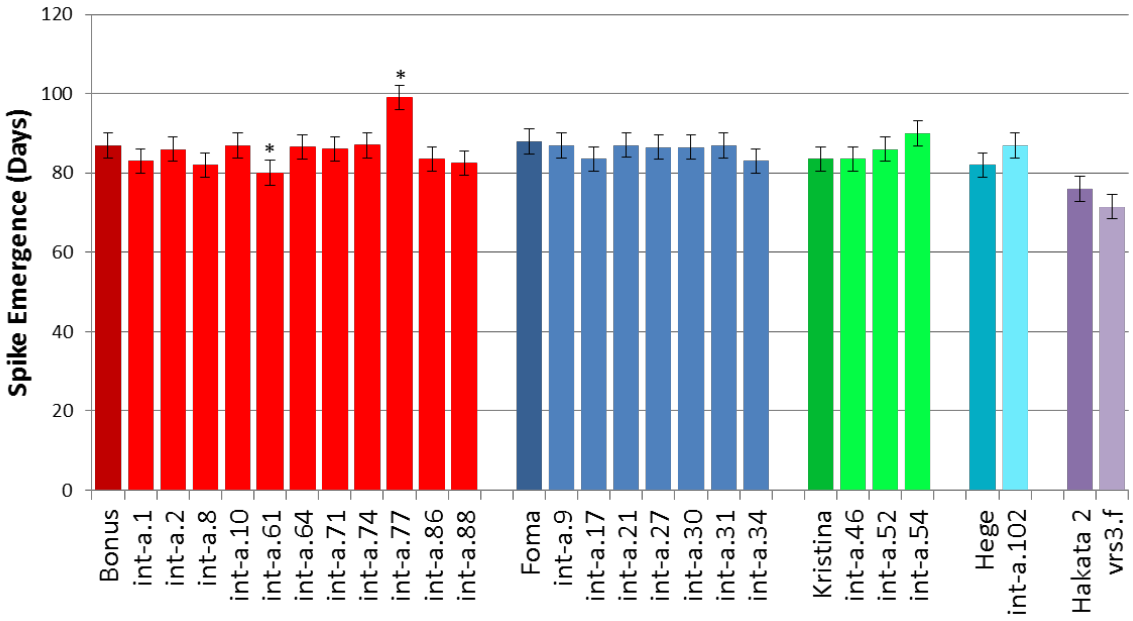


Figure 33 Comparison of spike emergence across the *vrs3* allelic series
*Colours represent alleles in the same genetic background with the parental genotype highlighted in a darker shade. Error bars show +/- SED. * represents those alleles with mean significantly different from their respective parental genotype.*

The majority of mutants showed no significant difference in spike emergence from their respective wild-type parents suggesting mutation of *Vrs3* does not impact spike-emergence (Figure 33).

The two alleles which did show a significant difference *int-a*.61 and *int-a*.77 were both from a Bonus background. Earlier cluster analysis (Figure 24) showed that whilst *int-a*.61 did cluster with the Bonus alleles, it was the most distantly related of the alleles; this could suggest that background genetic variability could underly the difference in spike emergence observed between *int-a*.61 and Bonus. This is also supported by the finding that the identical *vrs3* mutation in a Foma background, allele *int-a*.31, showed no significant difference in spike emergence compared to Foma.

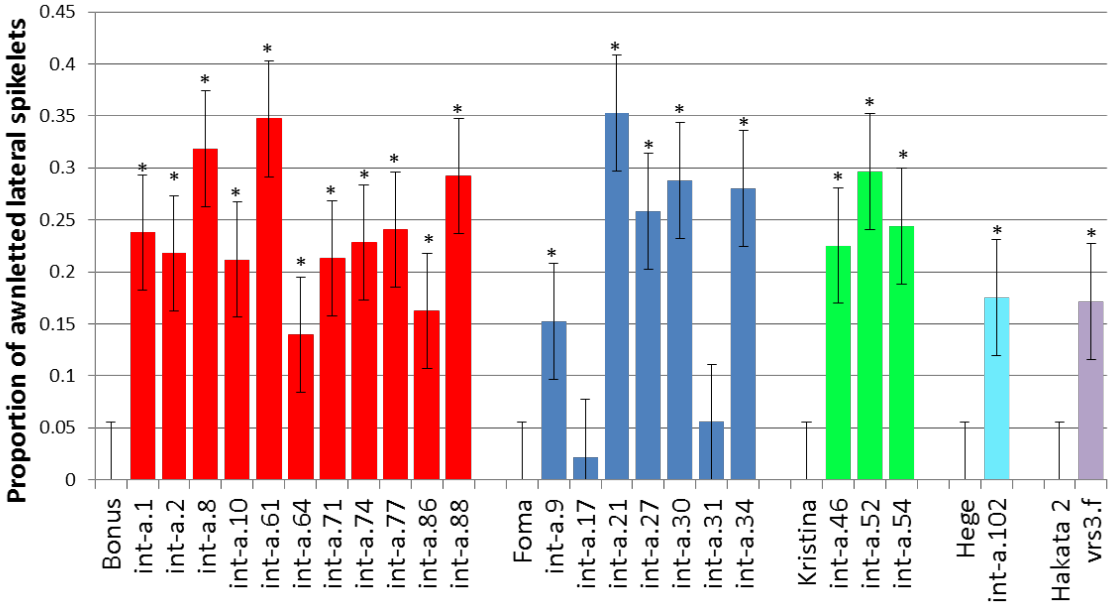
The *int-a*.77 mutant emerged significantly later than all other alleles, 99 days after sowing, 12 days later than Bonus. In the early stages of development GS10—GS25, both replicates of *int-a*.77 were observed to have a scorched leaf phenotype not observed in any other accession (Figure 34). The exact cause of this scorched phenotype is unknown but it is likely to have reduced the photosynthetic capacity and almost certainly delayed the growth of the plant. This phenotype therefore may explain the delayed heading date of *int-a*.77 compared to the other Bonus *vrs3* accessions.

Figure 34: The scorched leaf phenotype observed in the *int-a*.77 allelic mutant.
Plants shown are at the third true leaf developmental stage (GS13). The scorched leaf phenotype is clearly visible as white patches on the green leaf.

Lateral spikelet awn development

Lateral awn development was measured on two different scales, of which one measured the proportion of lateral spikelets with lemma awns of less than 2.5 cm; these were classified as awnletted. The second scale measured those lateral spikelets with lemma awns of greater than 2.5 cm; these were classified as awned laterals.

a)



b)

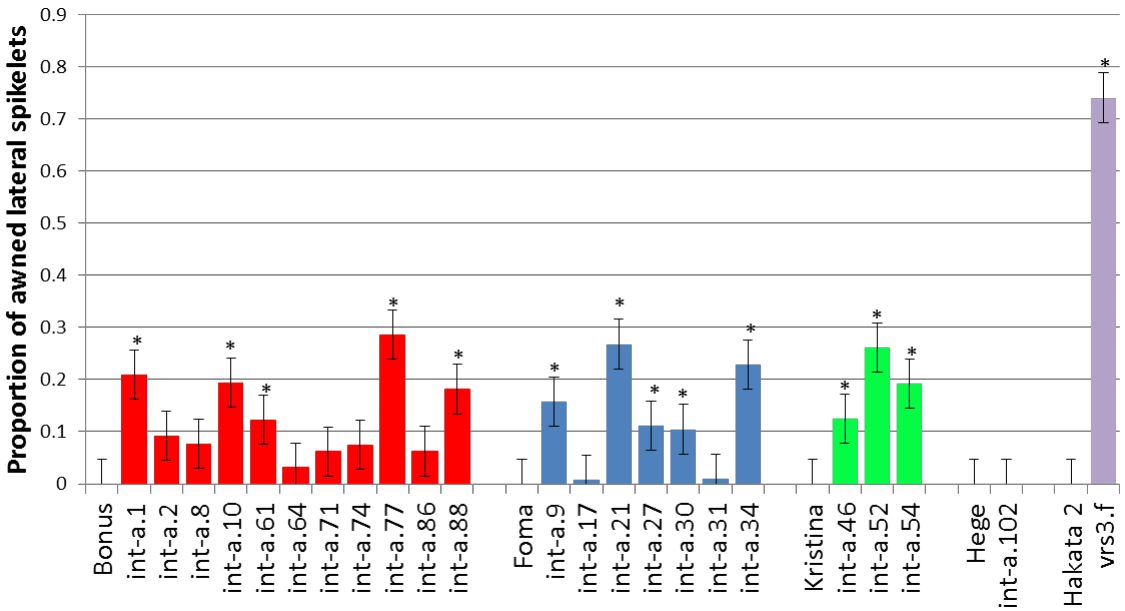


Figure 35: The proportions of a) awnletted and b) awned lateral spikelets across the vrs3 allelic series
*Colours represent alleles in the same genetic background with the parental genotype highlighted in a darker shade. Error bars show +/- SED. * represents those alleles with mean significantly different from their respective parental genotype.*

Across the domains of *Vrs3*, mutations outwith the three functional domains showed a significantly greater proportion of awnletted lateral spikelets ($p<0.001$). The awnletted

phenotype is considered a weaker mutant phenotype, intermediate between wild-type rounded and mutant fully awned. The fact that a greater proportion of awnletted laterals were observed in mutants outwith the functional domains of *vrs3* would fit with the idea of non-functional domain amino acids being more amenable to mutation, with less deleterious consequence than those mutations of the functional domains.

Awnletted lateral spikelets were most abundant in the *int-a*.21 mutant of Foma, with 35% of lateral spikelets awnletted, whilst *int-a*.17 had only 2% awnletted lateral spikelets (Figure 35). *int-a*.17 also showed no awned lateral development, suggesting it to be a relatively phenotypically weak mutant of *vrs3* i.e. more closely resembling wild-type two-rowed, this is despite it being a deletion mutant with the C-terminal of the JmjC domain and the ZnF domain deleted. A contrast between genetic background and mutant phenotype is observed again with respect to *int-a*.31 and *int-a*.61. Despite being the same nucleotide missense mutation, *int-a*.61 was found to have one of the largest proportions of awnletted lateral spikelets at 0.35 and *int-a*.31 one of the smallest at 0.05 (SED=0.04). This was also true of the awned spikelets, suggesting background factors are resulting in a stronger, more six-rowed, *vrs3* phenotype in *int-a*.61 compared to *int-a*.31.

The highest proportions of awned lateral spikelets were associated with mutations of the JmjN domain ($p<0.001$). Maximum awn development was observed in the *vrs3*.f mutant, with 74% of lateral spikelets awned (Figure 35). In contrast, spikelets of the *int-a*.17 and *int-a*.102 mutants showed no lateral awn development. Interestingly, *int-a*.71 (Bonus) and *int-a*.54 (Kristina), identical mutants in different genetic backgrounds showed significant difference in their lateral spikelet awn development; *int-a*.54 produced three times as many awned lateral spikelets compared to *int-a*.71. The reason for this is unclear but could reflect differences in the propensity of the two background cultivars to produce awns. Whilst no significant difference was found between the *vrs3* allelic series with respect to central spikelet awn length ($p=0.17$), there was a significant difference between the parent genetic backgrounds ($p<0.001$). Those lines in a Kristina background had significantly longer central awns than those lines in a Bonus background, 13.9 cm compared with 9.8 cm, (SED=1.2).

Lateral spikelet grain-fill

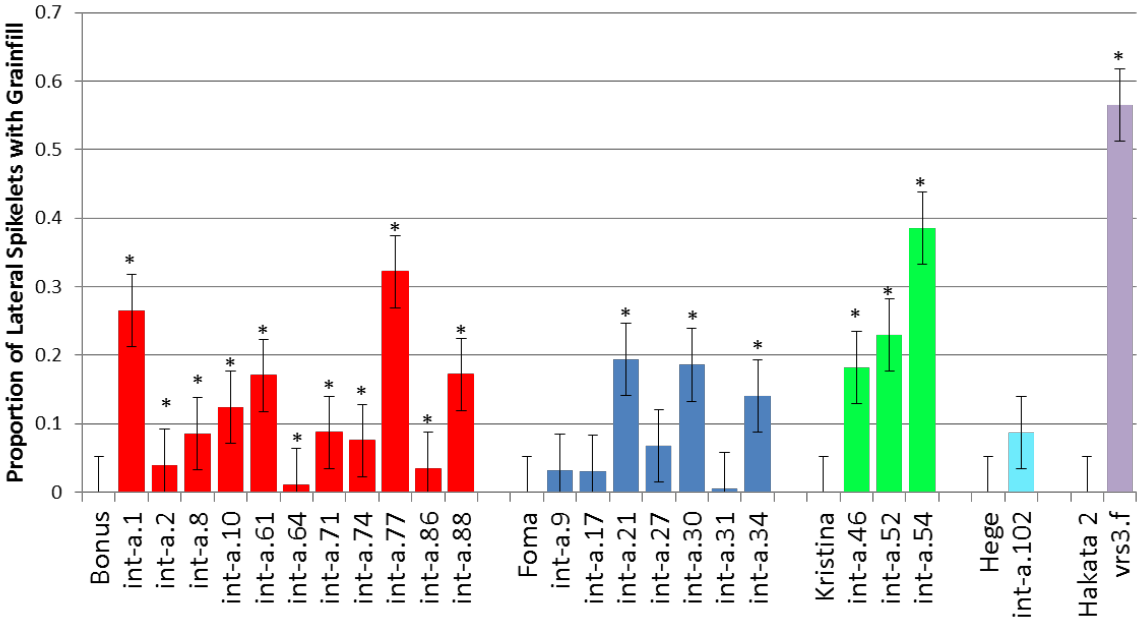


Figure 36: The proportion of lateral spikelet grain-fill across the accessions of the *vrs3* allelic series
*Colours represent alleles in the same genetic background with the parental genotype highlighted in a darker shade. Error bars show +/- SED. * represents those alleles with mean significantly different from their respective parental genotype.*

Mutations within the JmjN domain were associated with significantly higher levels of lateral spikelet grain-fill compared to mutations either within the JmjC, ZnF or outwith the functional domains ($p<0.001$).

vrs3.f showed the highest levels of lateral spikelet grain-fill at 56%, making this allele the most six-rowed spike phenotype (Figure 36). The lowest level of lateral spikelet grain-fill, 0.5%, was observed in the *int-a*.31 mutant. In contrast to *int-a*.31, the identical missense mutation in a Bonus background, *int-a*.61, resulted in 17% of lateral spikelets with grain-fill.

In addition to differences in genetic background influencing the *vrs3* phenotype, the type of mutation was also found to influence the lateral spikelet grain-fill phenotype. Foma mutants, *int-a*.30 and *int-a*.27, were identified as different non-synonymous mutants of the same *Vrs3* nucleotide. The G/T polymorphism in *int-a*.30 was predicted to result in a change from glycine to valine and the G/A polymorphism of *int-a*.27 glycine to aspartic acid. Comparison of lateral grain-fill between the two found a significant difference, with *int-a*.30 producing almost 3 times as much lateral grain than int-a.27, 18.5% compared to 6.8% respectively (SED=5.3%).

Awned palea

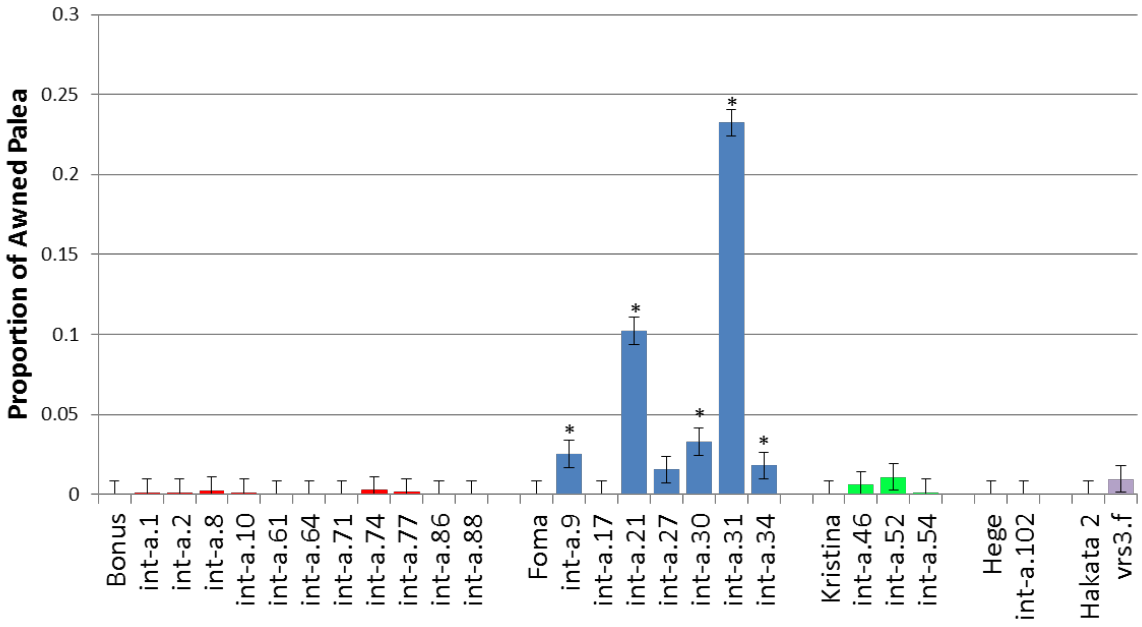


Figure 37: The proportion of spikelets with the awned palea phenotype across the vrs3 allelic series
*Colours represent alleles in the same genetic background with the parental genotype highlighted in a darker shade. Error bars show +/- SED. * represents those alleles with mean significantly different from their respective parental genotype.*

The awned palea phenotype was most associated with mutations in the JmjC domain and mutations outwith the functional domains ($p<0.001$). However, this finding may be confounded by the fact that the phenotype is largely associated with those *vrs3* mutants in the Foma genetic background, five of which are mutants of the JmjC domain.

The awned palea phenotype was less prevalent than the other scored phenotypes, with only two alleles with mean occurrence of over 10% (**Figure 37**). It also appeared to be highly genetic background specific; all Foma alleles, with the exception of *int-a*.17 and *int-a*.27 showed a significantly greater number of awned palea compared to wild-type ($p<0.001$). As with lateral spikelet grain-fill, the *int-a*.30 mutation at nucleotide 1769 gave a significantly stronger mutant phenotype than the *int-a*.27 mutation at the same nucleotide.

Tiller number

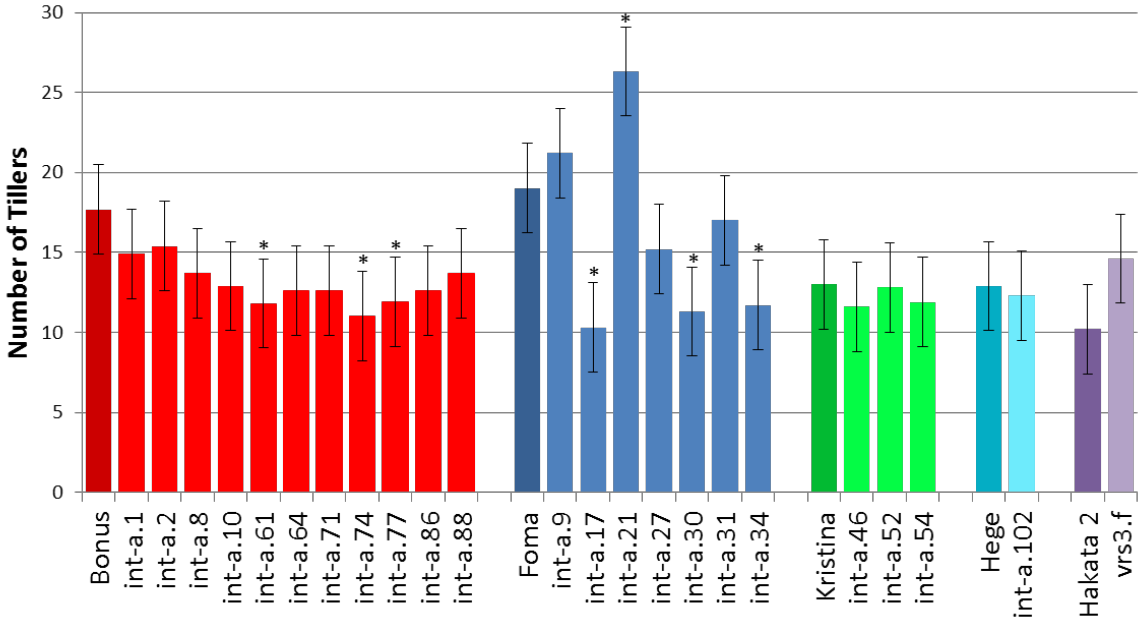


Figure 38: Mean tiller number at harvest across the *vrs3* allelic series
*Colours represent alleles in the same genetic background with the parental genotype highlighted in a darker shade. Error bars show +/- SED. * represents those alleles with mean significantly different from their respective parental genotype.*

Change in tiller number was not associated with mutation of one specific domain within *Vrs3* (p=0.278). A limited number of alleles showed significant difference in tiller number compared to their wild-type parent (Figure 38). *int-a*.17, *int-a*.30 and *int-a*.34 all produced significantly fewer tillers than Foma. In the Bonus background, *int-a*.74 and *int-a*.61 also produced significantly fewer tillers than Bonus.

int-a.21 produced the greatest number of tillers, 26 ,which was significantly greater than wild-type Foma (19, SED=2.789). Seed was scarce of *int-a*.21 so it was sown at a lower seed density compared to the other allelic series mutants this could explain its significantly greater tillering.

Supernumerary florets and spikelets

The supernumerary floret and spikelet phenotypes were not observed in all members of the *vrs3* allelic series, 8 alleles in both cases . Even within those alleles in which the phenotypes were observed, the occurrence was highly variable. Extra spikelets were identified in a maximum of 2 out of 10 spikes per allele. Extra florets were slightly more prevalent in some cases observed in up to five out of 10 spikes per allele.

The low numbers of observations made reliable statistical analysis of the allelic effects difficult. Moreover, a highly localised environmental interaction could be influencing the phenotype.

Amino acid conservation score

BLAST of the VRS3 predicted peptide sequence found orthologues of VRS3 to be highly conserved across plant species. To ascertain whether the mutant phenotypes identified are the result of mutations of highly conserved amino acid residues, a Shannon entropy conservation score was calculated for each of the amino acids within VRS3 based on their sequence alignment with 11 identified VRS3 orthologues. This measure of conservation takes into account whether amino acids at the same position in an alignment have conserved chemical properties[230].

Table 16: The amino acid conservation score of the residue mutated by each missense mutation in the *vrs3* allelic series.

Allele	Amino acid position	Conservation score
int-a.27	131	0.92202
int-a.30	131	0.92202
int-a.54	131	0.92202
int-a.71	131	0.92202
int-a.64	160	0.93233
int-a.46	229	0.88323
int-a.31	281	0.8618
int-a.61	281	0.8618
int-a.2	284	0.87151
int-a.9	301	0.89465
int-a.52	306	0.88435
int-a.10	399	0.99999
int-a.102	400	0.99999
int-a.86	531	0.64177

The highest conservation score of 0.99999 was found in 8 consecutive residues in the N terminal portion of the JmjC domain. Missense mutants of two of these residues are found in the *vrs3* allelic series, *int-a.*10and *int-a*.102. All the *vrs3* missense mutations bar *int-a*.86, have conservation scores greater than the *VRS3* median of 0.852 giving strong evidence that it is the mutation of conserved residues within *VRS3* that is giving rise to the phenotypes observed not random mutagenesis.

3.3.5.2 *Multivariate analysis of vrs3 phenotype*

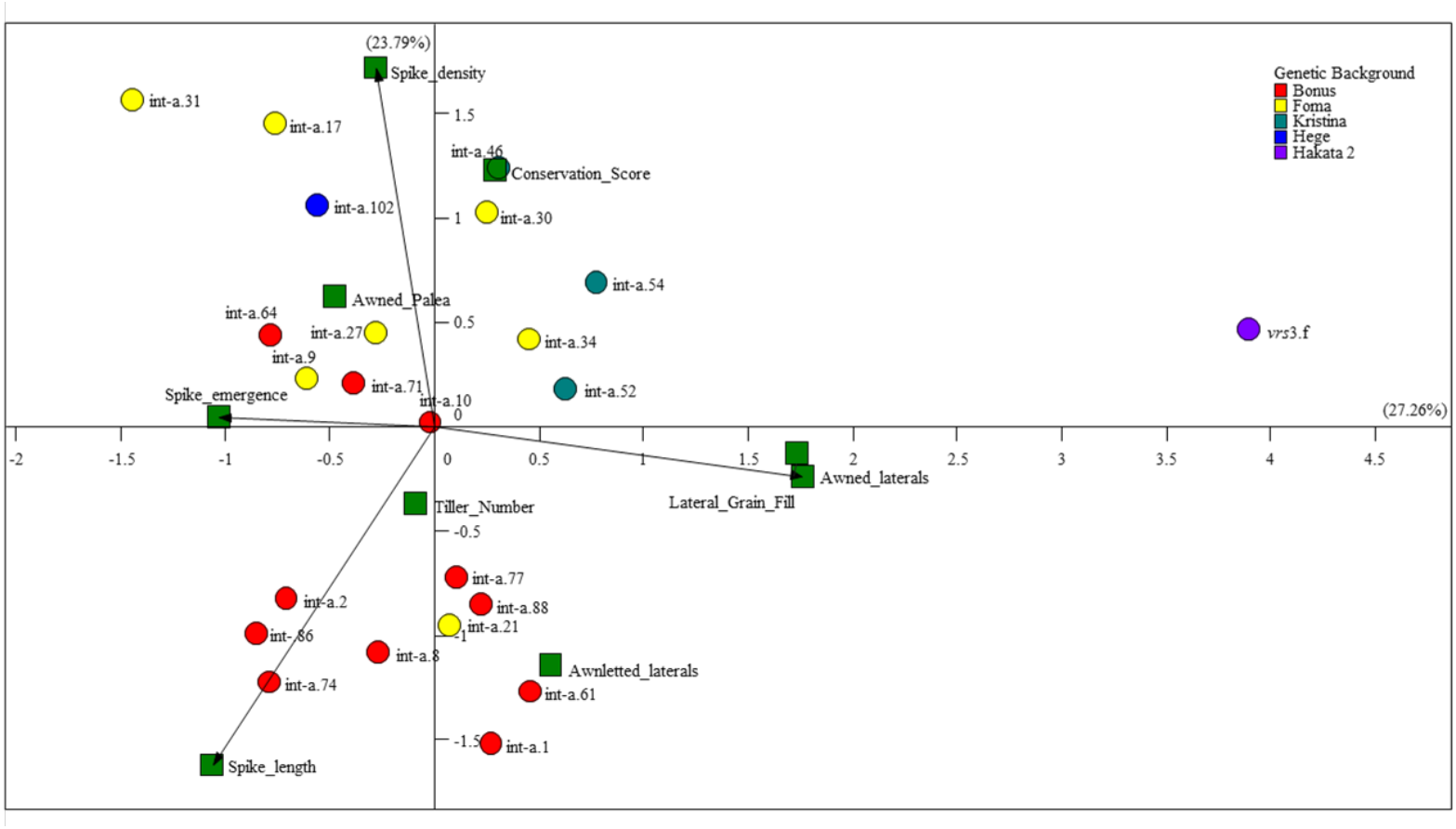


Figure 39: Multivariate analysis across the phenotypic means of the *vrs3* allelic series.
Individual vrs3 alleles are represented by coloured circles, with the colour of the circle denoting the genetic background of the allele. Phenotypic variates are represented by green squares.

Multivariate analysis (Figure 39) across the *vrs3* allelic series suggested that most of the variation in principal component 1 could be attributed to the spike emergence, lateral spikelet grain-fill and awned lateral spikelet phenotypes. The awned lateral spikelet and lateral grain-fill phenotypes were highly positively correlated but these were both negatively correlated with spike emergence, i.e. those alleles which took less time to reach spike emergence showed greater levels of lateral awn development and grain-fill. However, variation at these phenotypes is independent of variation associated with the amino-acid conservation score and spike-length, as shown by an approximate right-angle between the variates respective axes.

Variation at principal component 2 is mostly due to the spike density phenotype. This is highly correlated with the amino-acid conservation score and awned palea phenotypes. High levels of the awned palea phenotype were most associated with the alleles in a Foma genetic background, Foma alleles also had a denser spike; it is therefore not clear whether it is the same factor which relates the two, or genetic background effects.

The positive correlation between allele conservation score and spike density is interesting, with alleles which were due to mutation in highly conserved amino acids of VRS3 showing higher spike density, this could suggest a further role of *Vrs3* in control of spike density.

3.3.6 ***Vrs3 alleles in cultivated germplasm***

Investigation of the natural allelic diversity of *Vrs3* within cultivated germplasm can potentially provide insight into selection pressures from breeding and could provide a means of identifying any natural phenotypic diversity associated with the *Vrs3* locus. In the case of *VRS1* and *INT-C*, specific natural alleles have been identified which are wholly associated with one row-type. With respect to *VRS1*, *vrs1*.a1, *vrs1*.a2, *vrs1*.a3 and *vrs1*.a4 are all associated with six-rowed cultivated germplasm, with *Vrs1*.b2 and *Vrs1*.b3 associated with two-rowed cultivated germplasm. At the *INT-C* locus, whilst not able to produce a six-rowed phenotype alone, the *Int-c*.a allele is strongly associated with the *vrs1*.a alleles in six-rowed germplasm and the *int-c*.b allele with two-rowed cultivated germplasm.

3.3.6.1 *Vrs3 region initial haplotype identification*

To maximise the chance of resequencing as many cultivated alleles of *Vrs3* as possible, the Barley Genome Zipper[73] was used to identify SNP markers immediately flanking 11_21217. SNP Markers 12_11107 and 12_30786 were identified as flanking 11_21217, the SNP previously found to co-segregate with *VRS3* (section 2.3.4.2). Across the three SNP, six of the eight possible homozygous allele combinations were identified from the EXBARDIV cultivated germplasm collection (HVCC) [193] which had previously been genotyped with the 9K iSelect SNP chip[66]. 17 cultivars were selected for complete resequencing of *Vrs3* based on their growth-habit, row-type and allele combination in the vicinity of *Vrs3* (Table 17).

Resequencing of *VRS3* in the representative cultivars listed in Table 17 identified two predominant alleles of *VRS3*, characterised by six SNP in complete linkage disequilibrium (LD) with each other (Figure 24). To fit with international standard barley nomenclature, these alleles have been identified as *Vrs3.w* and *Vrs3.x* respectively. The positions of the SNPs distinguishing *Vrs3*.w and *Vrs3.x* are detailed in Table 18.

The SNP within exon 2 results in a Ser-Asn amino-acid change. Although non-synonymous, both amino acids are characterised by polar neutral side chains and therefore this SNP is unlikely to cause large conformational change within the protein, suggesting that both allele would result in functional proteins. Calculation of the Shannon entropy conservation score for both forms of the protein against 11 grass

orthologues of VRS3 found the VRS3.x form to show a greater level of conservation at this residue compared to the VRS3.w allele, 0.79454 compared to 0.76373.

In addition to the SNP listed in Table 18 a further SNP (G/T), was identified within the 5'UTR, -268bp from the start codon. This SNP further separated the *Vrs3*.w allele into two sub-alleles; those accessions with the G nucleotide at this polymorphism were identified by the allele *Vrs3*.w1 and those with the T, *Vrs3*.w2.

Table 17: Cultivars selected for their representative haplotype in the region of *Vrs3* and the allele of *Vrs3* identified on resequencing.

Cultivar	Growth Habit	Row-type	12_11107 (53.22)	11_21217 (54.73)	12_30786 (54.73)	*Vrs3* Allele
NFC Tipple	Spring	Two-rowed	A	A	C	*Vrs3*.w1
Edda	Spring	Six-rowed	A	A	G	*Vrs3*.w2
Abava	Spring	Two-rowed	A	A	G	*Vrs3*.w2
Bowman	Spring	Two-rowed	G	A	C	*Vrs3*.w1
Pearl	Winter	Two-rowed	G	A	C	*Vrs3*.w1
Olli	Spring	Six-rowed	G	A	G	*Vrs3*.x
Hatif de Grignon	Winter	Six-rowed	G	A	G	*Vrs3*.x
Lysiba	Spring	Two-rowed	G	A	G	*Vrs3*.w2
Malwinta	Winter	Two-rowed	G	A	G	*Vrs3*.w1
Steptoe	Spring	Six-rowed	G	G	C	*Vrs3*.x
Hoppel	Winter	Six-rowed	G	G	C	*Vrs3*.x
Orza	Spring	Two-rowed	G	G	C	*Vrs3*.x
Heligan	Winter	Two-rowed	G	G	C	*Vrs3*.x
Morex	Spring	Six-rowed	G	G	G	*Vrs3*.x
Bonus	Spring	Two-rowed	G	G	G	*Vrs3*.x
Tiffany	Winter	Two-rowed	G	G	G	*Vrs3*.x
Balda	Winter	Six-rowed	G	G	G	*Vrs3*.w2

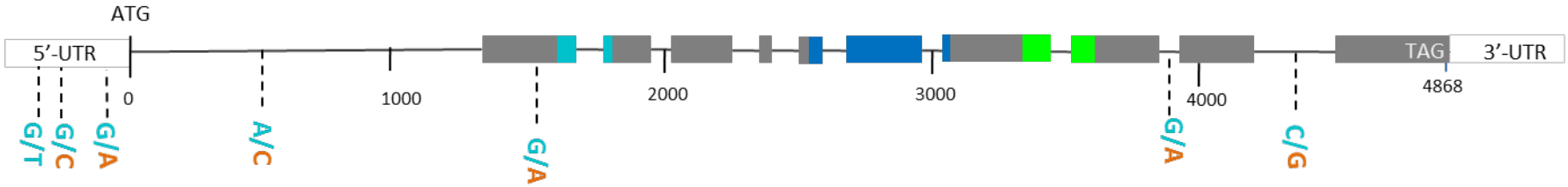


Figure 40: The six SNP distinguishing the *Vrs3*.w (turquoise) and *Vrs3*.x (Orange) alleles of *VRS3*.

Exons are represented in dark grey; the predicted functional domains are illustrated, JmjN-turquoise, JmjC-dark blue, ZnF-green. The additional SNP in the 5' UTR which divides Vrs3.w into Vrs3.w1 (G) and Vrs3w.2 (T) is also shown.

Table 18: The positions of the polymorphisms distinguishing *Vrs3.w* and *Vrs3*.x alleles

Polymorphism (*Vrs3*.w/*Vrs3*.x)	Nucleotide position relative to transcriptional start codon	
G/C	-276	5' UTR
G/A	-239	5'UTR
A/C	525	Intron 1
G/A	1583	Exon 2
G/A	3822	Intron 9
C/G	4325	Intron 10

It is apparent from the sequencing of *Vrs3* in representative cultivars that no one *Vrs3* allele is wholly associated with a particular growth habit or row-type (**Table 17**). Although, the *Vrs3*.w_1 allele has only been identified in two-rowed individuals in this small European cultivar subset, a larger germplasm analysis would be required to confirm whether this association is true.

3.3.6.2 *Allelic diversity across a larger germplasm pool*

To establish the relative proportions of the *Vrs3.w* and *Vrs3.x* alleles within cultivated germplasm, 408 individuals representing a diverse collection of largely European barley germplasm were genotyped with a KASP assay designed to the G/A SNP within intron 9.

Figure 41 shows the relative proportions of each allele across the winter and spring, two-rowed and six-rowed groups. Across the two-rowed winter germplasm, the proportions of *Vrs3.x* and *Vrs3.w* are equal. In contrast, there is preponderance of *Vrs3.w*, within the two-rowed spring germplasm. This would suggest a strong selection pressure either for *Vrs3.w* itself or other closely linked regions within the spring but not winter germplasm. Alternatively, the selection pressure exerted on the spring two-rowed germplasm could be present in the winters but balanced by a second selection pressure in the opposite direction.

Both *Vrs3.w* and *Vrs3.x* were represented within the genotyped six-rowed winter and spring germplasm. However, opposite alleles of *Vrs3* pre-dominated, *Vrs3.w* in the six-rowed springs and *Vrs3.x* in the six-rowed winters. Of the 23 spring six-rowed cultivars only 3 had the *Vrs3.x* allele; these were the North American cultivars, Steptoe and Morex and the German cultivar Frisia, which has winter six-rowed cultivar in its pedigree. A large proportion of the six-rowed spring cultivars originated from Scandinavia; this limited gene-pool could explain the predominance of *Vrs3.w* allele observed. The reason for the predominance of the *Vrs3.x* allele in the winter barley six-rowed germplasm is unclear, though the frost tolerance locus $FrH3^{231}$ has been associated with the region, and perhaps selection for this could underlie the allele skew observed. A list of all the cultivars, their *Vrs3* genotype and haplotype can be found in Appendix 17.

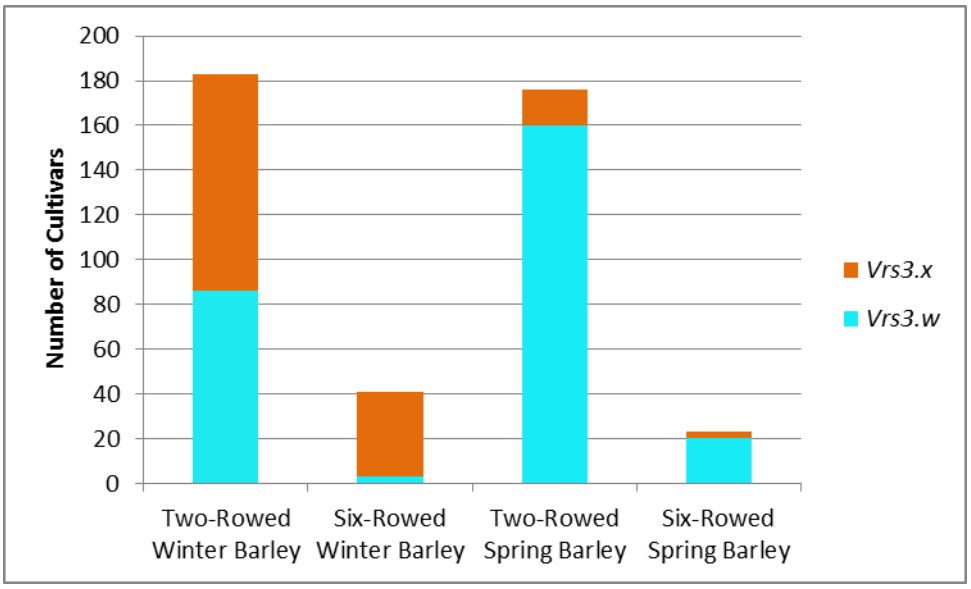


Figure 41: The proportions of the *Vrs3.w* and *Vrs3.x* alleles identified within cultivated germplasm.

Germplasm are separated based on their differing growth habit and row-types. Within each group the numbers of cultivars containing the Vrs3.x allele are illustrated in orange and the number of cultivars with the Vrs3.w allele illustrated in turquoise.

3.3.7 *Wider Vrs3 haplotype identification*

Following initial cultivar genotyping better consensus marker order in the region surrounding *Vrs3* was obtained from the IBGSC consensus SNP map[57]. Haploview was used to establish blocks of LD across chromosome 1H from iSelect SNP in IBGSC map order and thus enabled the identification of wider haplotypes in the region of *Vrs3*. *Vrs3* was identified as part of a block of 21 highly linked SNP (D'= 0.93). Haplotype analysis of this linkage block across the 408 cultivars previously genotyped for the intron 9 polymorphism identified 8 unique haplotypes with frequencies greater than 1% (Table 19). Of these, haplotypes 1 and 2 were considered major haplotypes, representing 50% and 25% of the haplotypic diversity respectively. The remaining six each accounted for less than 5% of the haplotype diversity within the population.

Table 19: The eight SNP haplotypes surrounding *Vrs3*.

The haplotypes highlighted in grey are associated with the Vrs3.x allele, those haplotypes not highlighted are associated with the Vrs3.w allele. SNP are ordered according to IBGSC consensus map marker position.

IBGSC Map position (cM)	48.58	48.69	48.8	48.8	48.8	48.8	48.8	48.8	48.8	48.8	48.88	48.94	48.94	48.94	48.94	49.17	49.17	49.26	49.75	49.75	50.05	
Haplotype	11_20810	11_11326	SCRI_RS_148600	11_10933	12_30750	SCRI_RS_14834	12_30350	SCRI_RS_152642	SCRI_RS_204810	SCRI_RS_221609	12_30110	11_20798	11_10520	12_30406	12_30694	Vrs3	11_21217	SCRI_RS_109060	12_30478	12_30499	11_21361	Proportion of population represented
1	A	G	G	C	A	A	G	G	T	G	A	A	A	A	A	G	A	T	G	G	G	0.5
2	G	A	A	G	G	T	A	A	C	T	G	T	G	G	G	A	G	C	A	A	C	0.25
3	G	A	A	C	G	A	A	A	C	T	G	T	G	G	G	A	G	C	A	A	C	0.047
4	G	G	A	G	G	T	A	A	C	T	G	T	G	G	G	G	A	T	A	A	C	0.042
5	G	G	A	G	G	A	G	G	T	G	G	T	G	G	G	A	A	T	A	A	C	0.034
6	G	A	A	G	G	T	G	G	T	G	G	T	G	G	G	A	G	C	A	A	C	0.025
7	A	G	A	G	G	A	G	G	T	G	A	A	A	A	A	G	A	T	A	A	C	0.02
8	A	G	A	G	G	T	G	G	T	G	G	T	G	G	G	G	A	T	A	A	C	0.017

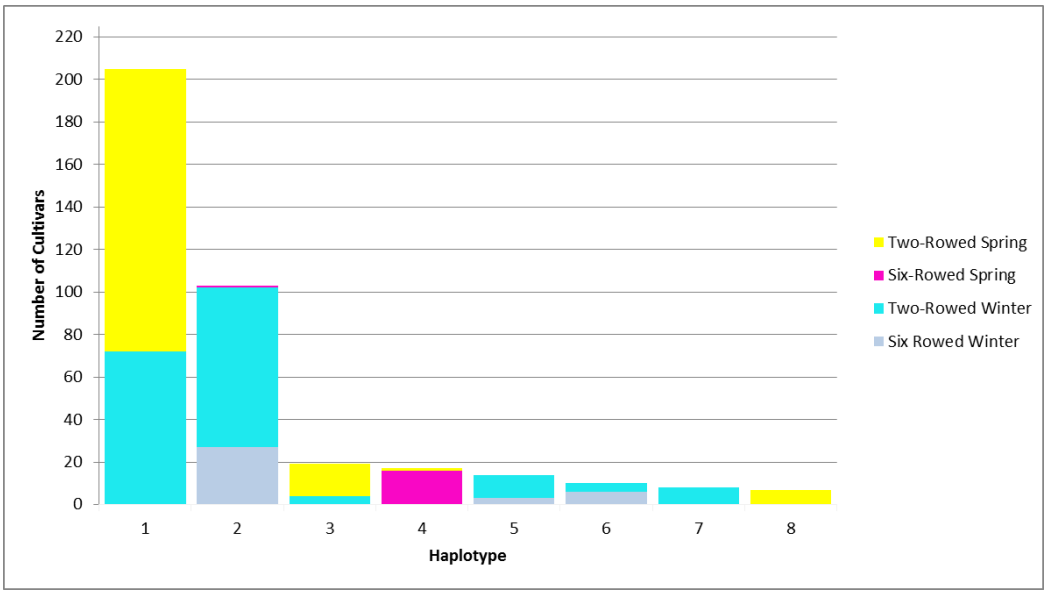


Figure 42: *Vrs3* haplotype diversity
The graph illustrates the numbers of cultivar represented by each haplotype divided into their respective growth habits and row-types. Haplotypes 1, 4, 7 and 8 are associated with the Vrs3.w allele and haplotypes 2, 3, 5 and 6 with the Vrs3.x allele.

Comparison of the types of cultivar across the identified *Vrs3* haplotypes (Figure 42) found 6 haplotypes associated with the winter two-rowed germplasm compared to 4 for the spring two-rowed germplasm. The larger number of haplotypes associated with the winter germplasm suggest either less selection pressure in the region within winter cultivars or selection for a variety of different traits, resulting in diverse haplotypes.

The majority haplotype, haplotype 1, comprises wholly two-rowed germplasm of both winter and spring growth habit. The germplasm within this haplotype are varied including Proctor, a UK spring cultivar first introduced in 1953 and Quench, a more recent UK spring cultivar, introduced in 2004 and a parent of many of the current UK (Recommended List) cultivars. The maintenance of the haplotype over time suggests either a low-level of recombination in the region, although there was no evidence for this in the F2 mapping of VRS3 or alternatively direct selection for a beneficial phenotype in the region.

The four two-rowed winter cultivars within haplotype 3, Igri, Willow, Magnolia and Nocturne are all linked by pedigree. Igri, a German cultivar, is a grandparent of the UK cultivar Willow, Willow is a parent of Magnolia and Magnolia a parent of Nocturne. Despite being distinct by 5 generations and four recombination events Igri and Nocturne still share the same haplotype suggesting very little recombination or strong selection for maintenance of the haplotype in this region of chromosome 1H.

Haplotype 4 was predominated by Scandinavian spring six-rowed cultivars confirming the earlier suggestion that the predominance of *Vrs3*.w observed within the six-rowed spring germplasm could be due to the limited gene-pool in which *Vrs3* was resequenced. The spring two-rowed cultivar within this class is Inari, a Finnish two-rowed cultivar with a six-rowed parent in its ancestry. The only other six-rowed spring cultivar within one of the eight identified haplotypes was the North American cultivar Morex in haplotype 2. Unlike the Scandinavian spring six-rowed, Morex has the *Vrs3*.x allele. Of the remaining 5 spring six-rowed cultivars genotyped, four unique haplotypes were identified (data not shown), which suggests that diversity in this region is present within the spring six-rowed germplasm but primarily outwith the majority of spring six-rowed population genotyped in this study.

From the re-sequencing of *Vrs3* in representative cultivars (Table 18) it was possible to infer haplotype *Vrs3* sub-allele associations. This suggested that haplotype 1 is the *Vrs3*.w1 allele and haplotypes 4 and 8 the *Vrs3*.w2 allele although further genotyping and/or sequencing would be needed to confirm this and the sub-allele of haplotype 7.

3.3.8 *Vrs3 and genome wide association with row-type*

Previous genome wide association studies (GWAS) for row-type, using both cultivated and landrace barley, have identified peaks on chromosome 1H in the vicinity of the *VRS3* locus. Following the identification of the gene underlying *VRS3* it was possible to test if previously identified GWAS peaks corresponded to allelic diversity at *Vrs3*.

The panel used to test the association was a subset of those cultivars included in the haplotype analysis (Appendix 17). PCoA (Principal coordinate analysis) confirmed the two-rowed and six-rowed individuals were separate distinct populations (Figure 43).

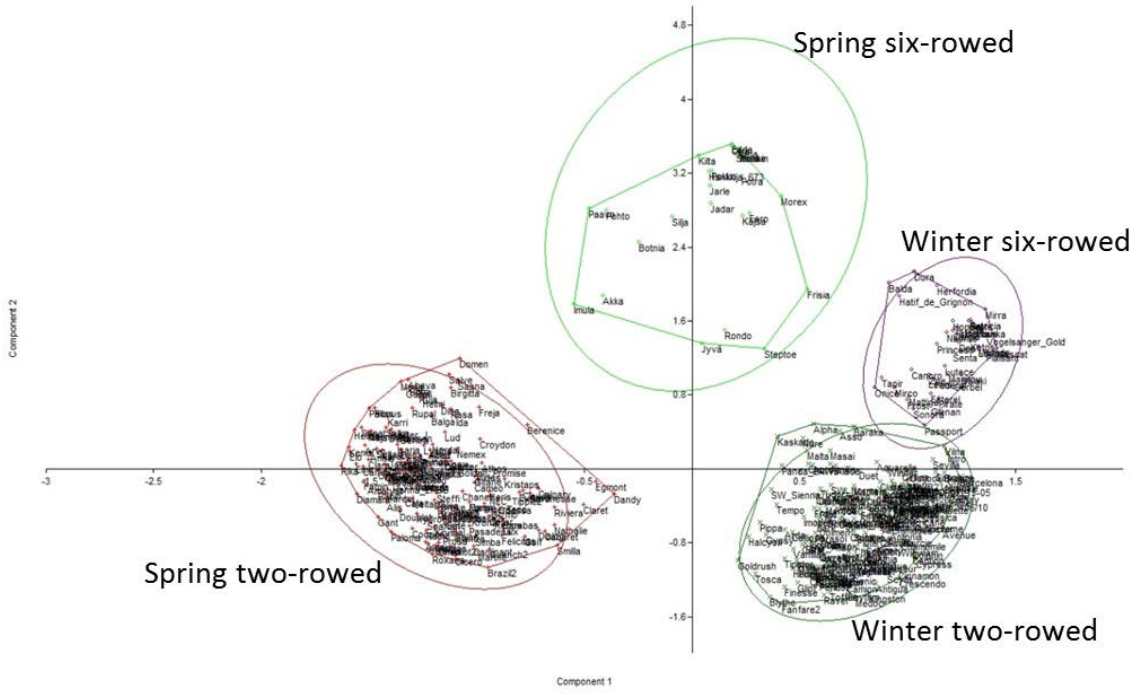


Figure 43: PCoA analysis illustrating the distinct germplasm populations used in the row-type GWAS study.
Spring two-rowed cultivars are bounded in red, spring six-rowed cultivars are bounded in pale green, winter two-rowed cultivars are bounded in dark-green and winter six-rowed cultivars are bounded in purple.

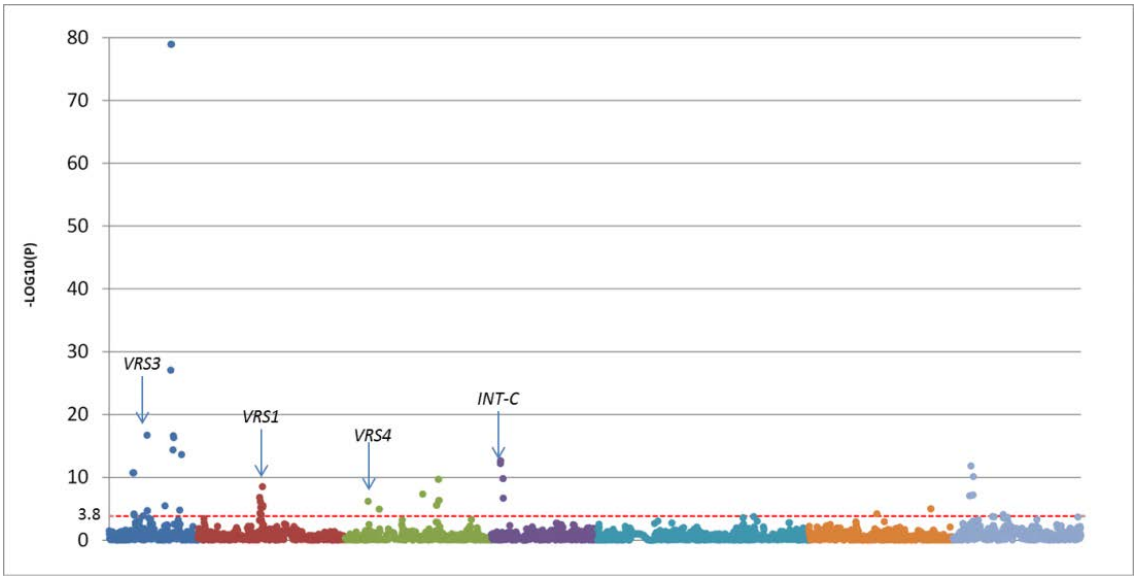
Two-rowed versus six-rowed represents a major source of structure within a barley germplasm population however it is unavoidable in a GWAS of row-type. To minimise other confounding factors of population structure, winter and spring germplasm populations were analysed separately. In all cases both a null association model, with no correction for population sub-structure and an Eigenstrat-analysis, including Eigen-

vectors within the association model, were fitted. This enables comparison between the two-models and aids in the identification of false positive associations.

Analysis of the spring association panel tested 4123 SNP against 159 two-rowed and 23 six-rowed cultivars. Null analyses produced a large number of significant SNP associations, 3172 at a False Discovery Rate of p=0.01. The Eigen-analysis produced much fewer significant marker row-type associations, 51 in total that effectively resolve into 9 QTL (Figure 44).

The winter association panel comprised 220 cultivars, 179 two-rowed and 41 six-rowed. 3138 SNP were included in the association analysis, which is 985 less than the spring panel and reflects the fact that the majority of the SNP on the iSelect panel were identified in spring germplasm; there is therefore a degree of ascertainment bias (non-random selection of SNP). Null analysis produced 2167 significant associations between row-type and SNP. Eigen-analysis produced 25 significant associations that effectively resolve into 7 QTL (Figure 44).

a) Spring barley row-type GWAS



b) Winter barley row-type GWAS

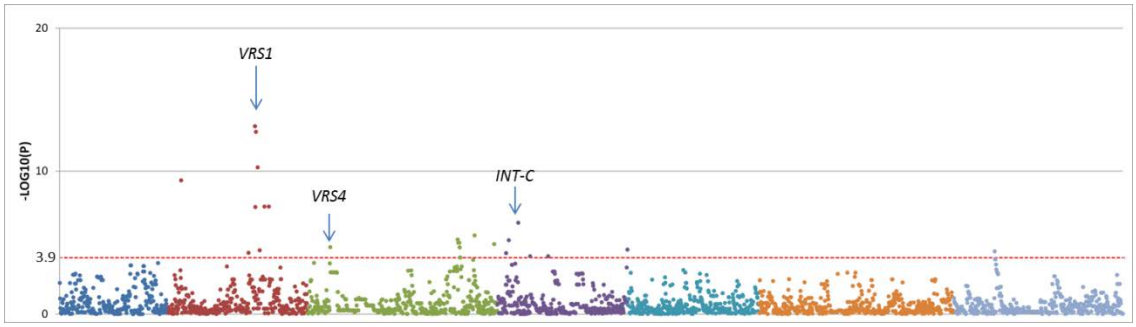


Figure 44a:Eigenstrat genome wide association scan of row-type across the seven chromosomes of a) spring barley and b) winter barley.
The Y axis shows –Log$_{10}$ (P) of the marker association probability values. Markers are plotted in relative order across the chromosomes. Chromosomes are shown in order from 1-7 with each identified by a different colour. The red-dashed line indicates a false discovery rate of 0.01. The significant peaks of previously identified row-type genes are highlighted.

Across both the winter and spring GWAS with structure correction, peaks were identified in the regions of previously identified row-type genes, *Vrs1*, *Int-c* and *Vrs4* (Appendix 18), giving confidence in the scan to detect genuine SNP row-type associations.

In both the spring and the winter GWAS the SNP representing the *Vrs3* allele was not found to have a significant association with row-type; this is as would be expected

given that the *Vrs3*.w and *Vrs3*.x alleles were identified in both two-rowed and six-rowed germplasm.

Within the spring GWAS the markers surrounding *Vrs3* were found to be significantly associated with row-type, with the peak marker SNP 12_10198, 2.32 cM from *Vrs3*. This follows with the haplotypes identified earlier, with the spring six-rowed primarily of one haplotype but the spring two-rowed sharing multiple haplotypes in the region. It is not clear whether a second peak in the region of Vrs3, SNP SCRI_RS_151764 at 46.71 cM is the persistence of LD from *Vrs3* or a separate locus. Earlier studies have mapped other spike morphology phenotypes including small lateral spikelets and lateral spikelet angle to the region of the second peak[184] which could suggest the peak is underlying a separate locus.

Despite the predominance of one allele within the six-rowed winter germplasm, a significant peak was not identified in the region of *Vrs3* within the winter association scan. This would be expected as both two-rowed and six-rowed individuals share a number of haplotypes, (haplotypes 2, 5 and 6) within the region therefore any marker row-type associations would be diluted. This highlights the fact that the detection of significant associations using GWAS is restricted to major alleles; in this case those with an allele frequency of greater than 10% in the population. The identification of significant peaks in regions where no known row-type loci have previously been mapped, e.g. on the long arm of chromosome 1H, suggests the presence of additional row-type factors within cultivated germplasm which remain yet to be identified.

3.3.9 *VRS3 diversity across plant species*

JmjC histone demethylases have been reported in many species both plant and animal[202,232-236]. Despite operating in diverse functional pathways the demethylase reaction which the JmjC demethylases catalyse appears conserved[224,226,237]. Identification of orthologues of VRS3 in other plant species could provide further information as to the metabolic pathways within which VRS3 functions.

Utilising reciprocal BLAST, 35 sequences from 23 plant species were identified as orthologous to VRS3. Figure 45 illustrates the maximum likelihood phylogenetic relationship of orthologues of VRS3 across the plant species.

The second JmjC barley locus on chromosome 6H. MLOC_53868.1 shares 50% overall protein sequence identity with VRS3. Focussing on the three functional domains protein identity increases to 70% at the JmjN domain, 85% at the JmJC domain and 57% at the ZnF domain.

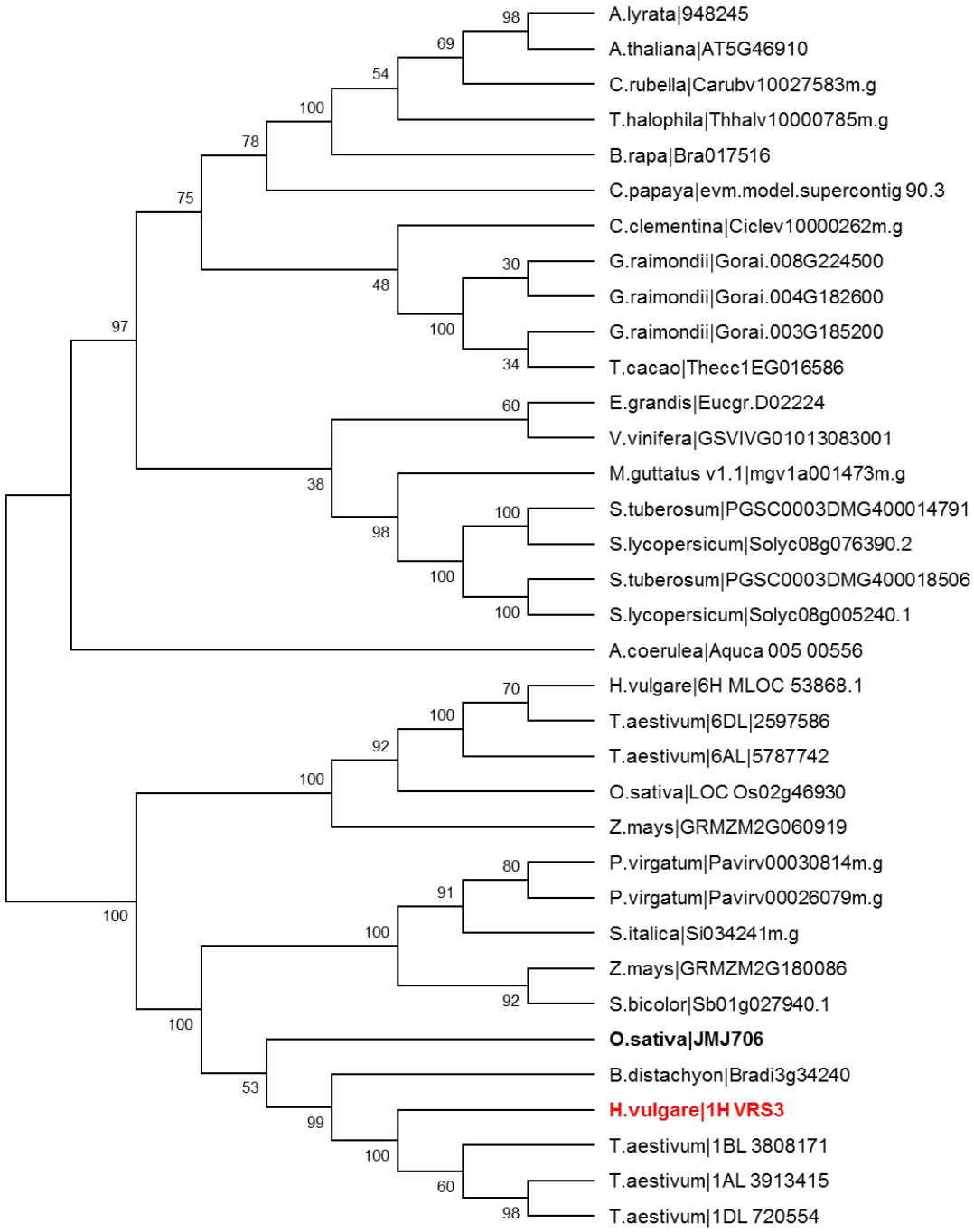


Figure 45: Maximum likelihood phylogenetic tree illustrating the relationship of VRS3 across plant species.

100 bootstrap replicates were used to infer the evolutionary relationships, the VRS3 *locus is highlighted in red and the rice orthologue JMJ706 in bold.*

3.4 *Discussion*

Despite morphological mutants of the *VRS3* locus being first classified over 60 years ago, it is only now that we have sufficient barley sequence data available to say with certainty that the gene underlying the *VRS3* locus is a JmjC histone demethylase.

3.4.1 *JmjC histone demethylases*

JmjC histone demethylases function within the cell nucleus to target specific lysine residues within the C-terminal tail sequence of histone proteins. Histone proteins are comprised of eight peptide subunits and act as a scaffold around which 147 DNA nucleotides wrap[238] (Figure 46). Differential chemical modification of specific residues within the histone tail sequence control the accessibility of regions of DNA to transcription with some chemical modifications associated with active transcription and others with repression[239].

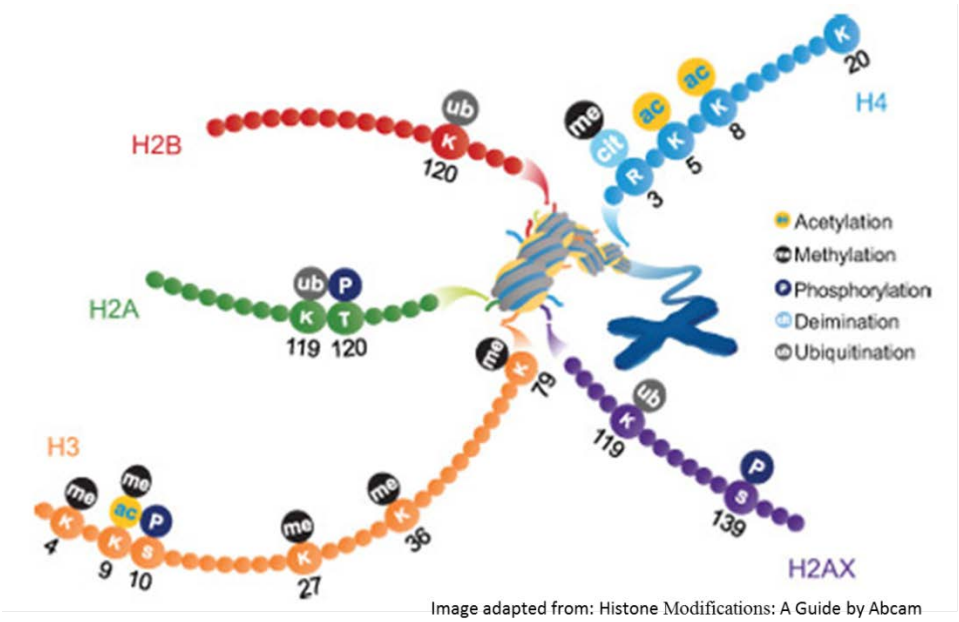


Figure 46: Cartoon representation of chromatin, histones and histone tail modifications. *Histones are represented by yellow spheres, around which grey DNA sequence is wrapped. Protruding from the histones are the histone subunit tail sequences, labelled according to the histone subunit to which they are attached. Highlighted in the tail sequences are the amino acids which undergo differential chemical modification. The chemical modifications are represented in circles above the residues of the histone tail sequence.*

Generally, JmjC histone demethylases remove methyl groups from lysine residues within the histone tail sequence[148,226]. However, the specific histone subunits and tail sequence residues targeted vary between JmjC demethylases[148] and the means of

targeting is not fully understood but studies have suggested that the histone tail sequence surrounding the lysine residue could influence specificity[240,241]

3.4.1.1 *JMJD2 histone demethylases*

Phylogenetic analysis of JmjC domain containing proteins across humans identified seven families, primarily differentiated by different functional domains in addition to the JmjC domain[148].

VRS3 was identified through synteny to the rice histone demethylase JMJ706. Sun *et al.*[202] identified JMJ706 to fall within the same clade as the human JMJD2 class of histone demethylases. The sequence homology between *JMJ706* and *VRS3* would suggest that *VRS3* is likely to be part of this clade too.

Human JMJD2 histone demethylases have been studied in some detail and implicated in the formation of a number of types of cancer[242].Their targets include the removal of repressive marks H3K9me2 and me3, H3K36me2 and me3 and H1.4K26me3. Investigations into the demethylase targets of JMJ706 found it to demethylate both H3K9me2 and H3K9me3 residues, with its mutation resulting in floral development defects[202]. However, although sharing JmjN and JmjC domains, the C terminal functional domains differ between human JMJD2, rice JMJ706 and barley VRS3. The C terminal of human JMJD2 features PHD and Tudor DNA binding domains[148], in contrast JMJ706 and VRS3 comprise a C5HC2 zinc finger DNA binding domain at the C terminus. Given this difference in functional domains between human JMJD2 and JMJ706/VRS3 further work is required before conclusions can be made as to the specific histone tail residue demethylation targets of VRS3. In addition, whilst histones are highly conserved across species, a recent paper in rice has identified species specific histone subunit variants[243], which may also have a role in determining the targets of the JmjC demethylases, again highlighting the need for further work prior to concluding the nature of VRS3 targets.

3.4.1.2 *JmjC histone demethylases in barley and other plants*

VRS3 is only the second JmjC histone demethylase to be identified in barley; the first *HvPKDM7-1* was identified as a putative H3K4 demethylase through orthology with the Arabidopsis gene *AtPKDMB*[244]. HvPKDM7-1 is located on the long arm of chromosome 1H towards the telomere, 131.2 cM (IBGSC map). In a comparison of two differentially drought tolerant cultivars Caresse and Demetra, *HvPKDM7-1* was found

to be significantly upregulated in the drought tolerant cultivar Demetra after 10 days of drought stress. However, the specific histone tail residues targeted are yet to be verified within barley and the specific genes which are differentially regulated under the induction of HvPKDM7-1 remain to be identified[244].

In addition to the *VRS3* orthologue *JMJ706*, a number of other JmjC histone demethylases have been identified in rice. These include: *JMJ703* associated with the demethylation of H3K4 residues. *jmj703* plants are characterised by reduced stem length attributed to increased levels ofH3K4 methylation in the histone subunit tail sequences associated with the promoter regions of cytokinin genes, resulting in reduction in their expression. Another gene *JMJ705* has been characterised as an H3K27me3 demethylase, which mediates a resistance response against rice bacterial blight. The rice induced early flowering mutant *se14* was mapped to chromosome 3 and identified as an H3K4me3 demethylase which targets the floral promotion locus *RFT1*[232].

Within Arabidopsis, JmjC histone demethylases have been characterised in a number of different genetic pathways including flowering time (*ELF6* and *REF6*), the circadian clock (*Jmj30*), salt tolerance (*Jmj15*)[234] and DNA methylation (*IBM1*). Surprisingly, despite mutant accessions being available[245], the *VRS3* orthologous locus of Arabidopsis (At5G46910) has not been characterised, suggesting perhaps a subtle phenotypic difference between mutant and wild-type or alternatively the genetic pathway within which it functions is yet to be characterised.

Phylogenetic analysis found orthologues of *VRS3* to be conserved across a wide variety of plant species; however the biological pathway within which *VRS3* functions is yet to be determined. Phenotypic evidence from analysis of the F_2 and F_3 mapping populations and the *vrs3* mutant allelic series suggests *VRS3* does not play a role in the flowering time pathway under long day conditions with no significant difference in spike emergence observed between mutant and wild-type. However, the genetic background, particularly with respect to spike emergence, is influencing the *vrs3* phenotype, as highlighted by multivariate analysis, with mutants which showed earlier spike emergence showing enhanced lateral spikelet grain-fill.

Growing the *vrs3* allelic series in a variety of environmental conditions, differential temperature and nutrient regimes may help to further ascertain the pathways within

which *VRS3* functions. Introgression of the *vrs3* allelic series into a common genetic background will help distinguish the phenotypic effects of genetic background compared to environmental factors. Phenotypes of the *vrs3* allelic series suggest a broad range of effects of the varying mutants, with some retaining a predominantly wild-type two-rowed phenotype and others a more mutant six-rowed phenotype. This and the fact that the mutations are not evenly distributed throughout the gene would suggest that the causal polymorphisms are disrupting key residues rather than entire domains within the gene.

The number and diversity of mutants identified within the *vrs3* allelic series represents a valuable resource to study the general function of JmjC histone demethylases in detail. However, given the effect of genetic background it may be more informative to form sets of near isogenic lines in diverse genetic backgrounds, winter/spring, two-rowed/six-rowed to assess the implications of the mutations fully.

3.4.2 ***Allele phenotypes***

The varying morphological phenotypes observed within the *vrs3* allelic series would suggest that genes underlying these phenotypes are potentially under the regulation of *Vrs3* either directly or indirectly. The awned palea phenotype, observed particularly in the *vrs3* mutants originating from Foma, has previously been reported as a spontaneous mutational event (*adp1*.a)[246]. The gene underlying the *adp1*.a locus remains yet to be determined, however, comparison of BOPA1 genotyping of the *adp1*.a Bowman near isogenic line (GSHO 1950) and Bowman found segregating markers spanning a 62 cM interval of chromosome 3H (11_10825 (38.03 cM) to 11_10312 (100.28 cM) (IBGSC map)) [114]. Investigation of the haplotypic diversity of Foma compared to the other *vrs3* mutant backgrounds, Bonus, Kristina and Hege may aid with targeting the gene underlying the *adp1*.a locus further, as given the prevalence of the phenotype within the Foma background, it would seem likely that it may have a novel haplotype in the region of *adp1*.a.

The correlation observed between lateral spikelet fertility and awn development has previously been observed at the barley *Lks2*[247] (short-awn 2) locus; the *lks2* locus is allelic to the *ubs4* [247] (*unbranched style-4*) locus. In addition to changes in awn length, mutation at this locus is associated with changes in pistil morphology and reduced fertility. An *SHI* (*Short Internodes*) transcription factor underlies this locus[248]. A further

18 loci have also been linked to awn development[249]; given the phenotype of *vrs3,* it would seem likely that some of these loci are regulatory targets of *Vrs3*.

The allele data suggest that mutation of *vrs3* results in a significantly laxer spike, compared to the wild-type parent. Within barley three mutant *laxatum* loci have been identified, *lax-a* (5H), *lax-b* (6H) and *lax-c* (6H)[113], of which one, *lax-a* , has been cloned[250], these loci could represent potential demethylation targets of *Vrs3*.

The strong negative correlation between spike emergence and the proportion of lateral spikelet fertility observed is interesting, with stronger mutant phenotypes observed with earlier spike emergence, yet no significant difference was observed between mutant and parent spike emergence. One interpretation of this could be that an as yet unidentified environmental cue is inducing a pathway which suppresses the effect of the *vrs3* mutant. However, the earlier emerging cultivars are at a later stage of inflorescence development so the suppressive effects are less in these cultivars than those which emerge later.

The two primary cues which have been implicated in barley spike-emergence are temperature, with inflorescence development suppressed prior to vernalisation in winter barley, and photoperiod. Currently all *vrs3* mutants are in the spring background and therefore carry deletions at the photoperiod sensitivity loci *ppdh1*[153] but functional allele at the *Ppdh2* locus [251,252]and also lack a vernalisation requirement. The *vrs3* alleles have only been grown under long day conditions thus far; however, assessment of allele phenotype under short-day conditions may establish any link between *vrs3* and the *Ppdh2* locus. With respect to vernalisation, an interaction between the row-type locus *VRS1* and vernalisation locus *VRN1* has been identified[253]. Introgression of the *vrs3* mutant phenotype into a winter background may help to further understand whether the well characterised photoperiod/vernalisation pathways influence the phenotype.

3.4.3 ***Diversity within cultivated germplasm***

The current study of diversity of *Vrs3* within cultivated germplasm has identified two major alleles, *Vrs3*.w and *Vrs3*.x, with *Vrs3*.w split into two sub-alleles *Vrs3*.w1 and *Vrs3*.w2. Resequencing of selected haplotype representatives followed by KASP assay genotyping of a single SNP provided a cost-effective means of genotyping large numbers of individuals. However, this approach does assume complete linkage between the genotyped SNP and *Vrs3* allele.

Mapping of *VRS3* (Section 2.3.3) positioned it centromeric on chromosome 1H; centromeric regions of the chromosome are associated with low levels of recombination[150] which could explain the persistence of haplotypes in the region of *Vrs3*. Alternatively, directed selection could underlie the observed haplotype retention; this theory is supported by previous mapping studies in which frost tolerance loci[193,231] and malting quality[200,254] attributes have been mapped to the same region of chromosome 1H as *Vrs3*. The selection for these traits could also explain the greater number of SNP haplotypes associated with *Vrs3* in the winter barley germplasm compared to the spring germplasm, with selection in the winter germplasm for both malting quality and frost hardiness resulting in multiple haplotypes whereas selection in the spring germplasm would be more focussed towards malting quality.

Whilst haplotype diversity is in evidence in the region surrounding *Vrs3* it does not persist within *Vrs3* itself, even individuals that represented minor haplotypes when re-sequenced retained one of the three alleles of *Vrs3*. This level of diversity seems relatively low especially when compared to the 17 alleles associated with *VRS1*[255]. This could suggest limited origins of *Vrs3* into cultivated germplasm. Genotyping of a wider globally representative collection would confirm this. Additionally, genotyping within landrace and wild barley species *H. spontaneum* would also help to confirm levels of diversity.

3.4.4 ***Association analysis***

The inclusion of the *Vrs3* allele specific SNP in the GWAS studies did not give a significant association between row-type and *Vrs3* allele. This is to be expected given the division of alleles across both the two-rowed and six-rowed germplasm if the natural variants are not essential for six-row formation. With the gene underlying *Vrs3* now identified it seems likely that rather than being associated directly with a change in row-type, *Vrs3* functions upstream in the row-type determination pathway, perhaps modulating the expression of other identified row-type genes. A similar relationship between row-type genes has already been identified with respect to the control of expression of *Vrs1* by *Vrs4*, with *Vrs1* expression highly down-regulated in *vrs4* mutants[130].

The significant association in the region of *Vrs3* identified in the spring germplasm is the result of the majority of six-rowed alleles representative of a single haplotype, haplotype 4. Whilst the association with row-type is true, the majority of spring six-

rowed have the *Vrs3*.x allele and the majority of spring two-rowed germplasm have the *Vrs3*.w allele; more work is required to ascertain whether it is specific selection for the *Vrs3*.w allele in haplotype 4 rather than an artefact of population structure. However, given the nature of barley breeding in Europe, the spring six-rowed market is very much restricted to Scandinavia[39,256], so although biased, the GWAS population used does represent a fair reflection of European spring six-rowed germplasm.

A row-type genome wide association study of the USDA barley core collection identified the SNP marker 11_10933 as the peak association for *Vrs3*[67]. Whilst found not to be significantly associated with row-type in this study, 11_10933 is within the identified *VRS3* haplotype block, with the C polymorphism associated with haplotypes 1 and 3 and the G polymorphism the remaining 6 haplotypes. However, neither polymorphism is in complete LD with the allele of *Vrs3*, hence the lack of a significant association suggesting that population used in the Barley Core collection[67] had an alternative population structure to that used in this GWAS study; to truly confirm the peak in the Barley core collection study, the collection would require genotyping for the specific allele of *Vrs3*.

3.4.5 ***Conservation across species***

In addition to highlighting VRS3 conservation across plant species, phylogenetic analysis has identified a duplicate of VRS3 on chromosome 6H. Duplicates were also identified in rice on chromosome Os2. The positions of the duplicates are consistent with a duplication event occurring in a common ancestor to the Poaceae rather than separate intra-specific duplications[257]. This supports the findings of a wider phylogenetic analysis of JmjC domain containing proteins across plants and animal which found duplications of this clade of protein within rice but not Arabidopsis[258].

Further work would be required to ascertain whether the paralogue of *Vrs3* is functional and if so what its function is. Although this is outwith the scope of this study it could provide a means of further understanding the role of JmjC histone demethylases within barley.

It is clear from Figure 45 that VRS3 is not confined to the monocots but found across a diverse range of eudicots plant species. Its presence in both monocots and eudicots suggests an early evolutionary origin[259].

Within the monocots, the VRS3 orthologues split into two major clades, reflecting the two orthologues in barley. One clade contained VRS3 and the second contained the chromosome 6H orthologue MLOC_53868.1. The chromosomal positions of the two loci within rice, barley, maize and wheat reflect the 5 ancestral chromosome model proposed by Salse *et al.*[260] suggesting the *Vrs3* duplication event to have occurred early in cereal evolution, likely during the whole genome duplication event, ~60Mya[257].

The presence of orthologues of VRS3 across a diverse range of plant species would suggest a general role perhaps with respect to floral morphology, given the mutants of rice and barley. However, despite the large number of orthologues within economically important crop species, prior to this study the only other species *HvJMJ706* has been characterised in is rice.

4. *VRS3* in Combination with *VRS1* and *INT-C*

4.1 *Introduction*

In the UK, six-rowed barley is restricted to the winter animal feed market. This is in contrast to other parts of Europe, particularly France, and North America where six-rowed barley is readily used for malt for the brewing industry. Historical perception associates six-rowed barley with low specific weight (the volumetric weight of grain expressed as kg/hl) [31,261], high levels of screenings[261] and uneven grain size[24,262] resulting in uneven modification of the grain into malt[8]. These negative perceptions stem from the fact that the lateral grain in six-rowed barley tend to be smaller than the central grain.[24,262] Previous studies of the central and lateral grain size of six-rowed cultivars have suggested differences of 30% with respect to grain weight[24,262]. However studies across a historical collection of six-rowed barley cultivars from the USA has shown improvement for grain plumpness[263,264], particularly lateral grain[263], with breeding.

Conventional cultivated six-rowed barley is comprised of the following natural alleles, *vrs1*.a and *Int-c*.a, at the unlinked loci *VRS1 (*chromosome 2H) and *INT-C* (chromosome 4H), with the *vrs1*.a allele on chromosome 2H sufficient to cause the six-rowed phenotype. The role of *Int-c*.a in the six-rowed phenotype remains unclear, it has previously been suggested it may have a role in promoting tillering[72] and enhanced lateral grain fertility[84,88].

Whilst crossing different induced intermedium alleles to determine allelism, Gustafsson *et al.* noted that some combinations produced a full regular six-rowed phenotype when in double recessive mutant combinations[265]. Following this a number of studies, primarily by Lundqvist *et al.*[94,110,132-134,266-270] have investigated the interaction of different induced intermedium mutant combinations on the row-type phenotype. This work was carried out prior to the identification of *HvHox1*[78], the gene underlying *VRS1,* and the *hex-v* and *int-d* alleles were presumed alleles of different tightly linked loci. Following work from Komatsuda *et al.*[78] we now know that the *int-d* and *hex-v* mutants are in fact all alleles of the *VRS1* locus. The *hex-v* mutants were classified as such due to a more six-rowed homozygous phenotype than the *int-d* mutants[265], also producing a phenotype of enlarged infertile pointed laterals in the heterozygous state[267]. This

difference in phenotype is presumably due to the type and position of the mutation within *HvHox1*.

The work of Lundqvist *et al.* focussed on three main phenotypes when categorising the mutant combinations: lateral spikelet awn length, lateral spikelet fertility and lateral spikelet seed size.

In the intermedium collections, the most numerous alleles are found in the *int-a (*allelic to *vrs3), int-c*, *int-d* (allelic to *vrs1*), *hex-v* (allelic to *vrs1*), and *int-e* (allelic to *vrs4*) loci, therefore the majority of work focussed on interactions between these loci. Of particular interest to this study of *vrs3* is the finding that *int-a* was found to combine well with *hex-v*/*int-d* and *int-c* to produce regular six-rowed phenotypes but in combination with *int-e,* more irregular spike phenotypes were observed[132,265].

Findings of Lundqvist *et al.* work can be summarised as follows:

A combination of two homozygous recessive intermedium loci are capable of producing a regular six-rowed phenotype[132].

For heterozygous intermedium loci to produce an enhancement in lateral spikelet phenotype, another intermedium locus has to be in present in homozygous recessive form[132].

The combination *hex-v* + two homozygous recessive intermedium loci produce a particularly enhanced six-rowed phenotype[133].

Varying the allele of a specific intermedium locus included within a cross can lead to heterogeneity with respect to the phenotype observed[132-134].

Environmental conditions can influence the phenotype observed, particularly with respect to the *int-c* locus[94].

Crosses combining the *hex-*v_A allele of six-rowed cultivar Agneta with intermedium alleles produced a stronger phenotype than the phenotypically strongest induced *hex-v* mutant *hex-*v_3[267].

Enhanced phenotypes were noted in three and four intermedium gene combinations but due to the lack of molecular markers, assignment of genotypes relied on the interpretation of phenotypes, with particular problems encountered when trying to

identify triple or quadruple homozygous recessive intermedium combinations from double mutant recessive combinations and heterozygotes[134].

Segregating genetic background may also have influenced the phenotypes observed as the intermedium mutants were induced in a range of two-rowed cultivars. Additionally, the non-targeted nature of the original mutagenesis means there is potential of additional mutational events within the genetic background.

Other research in barley focussing on grain size has identified QTL in the region of *VRS1* and *INT-C*[271] but no study has been able to conclusively genotype at three row-type loci and look at the specific contribution of each allele.

4.1.1 ***Chapter summary***

The aims of this chapter are to further understand the interaction between differing allele combinations of the row-type genes *VRS1*, *INT-C* and *VRS3* and their phenotypic effect particularly with respect to inflorescence architecture utilising populations from the bi-parental crosses Morex x .BW419 and Morex x BW902.

Initial phenotypic assessments aimed to ascertain if it was possible to identify different alleles of *VRS3* in the three-way row-type gene segregants. It also aimed to test what effect differing alleles at *VRS1* and *INT-C* had on overall spike architecture and tiller number in field and glasshouse conditions.

Subsequent identification of *VRS3* (Chapter 3) and definitive genotyping within the F2 populations enabled further detailed phenotypic assessments of central and lateral grain parameters. These were used to ascertain the effect of the varying genotypic combinations of *VRS1*, *INT-C* and *VRS3* on grain size and number and to test the hypothesis that the three-way homozygous six-rowed genotype (*vrs1*.a, *Int-c*.a, *vrs3*) does give a significantly better six-rowed phenotype with more uniform lateral grain-fill, when compared to the current cultivated six-row model of *vrs1*.a, *Int-c*.a, *Vrs3*.

4.2 *Materials and methods*

4.2.1 *Germplasm*

F2 populations were created as previously described in 2.2.1 between *vrs3* Bowman isolines BW902 (*vrs3*.f) and BW419 (*int-a*.1) and the six-rowed cultivar Morex.

Morex is a North-American six-rowed spring cultivar, bred by The University of Minnesota from the cross Cree/Bonanza and first registered in 1979[272]. Populations were identified as described Table 20.

Table 20 Population identifiers for the *vrs3* Bowman isoline by Morex crosses

Population Identifier	Cross
BW419Mo	Morex/BW419
BW902Mo	Morex/BW902

4.2.2 *F_2 population—BW902Mo and BW419Mo phenotyping*

4.2.2.1 *Experimental design*

Unlike the crosses to Bowman and Barke, the Morex F_2 populations have three row-type genes segregating (*VRS1, INT-C and VRS3*). Morex donated six-rowed allele's *vrs1*.a and *Int-c*.a, and wild-type *Vrs3* to the cross, whilst BW902 and BW419 donated two-rowed alleles *Vrs1*.b and *int-c*.b and mutant *vrs3*. To ensure a 99% chance of all 27 row-type gene combinations being represented in the population, a population size of 600 individuals, 300 from each cross was selected[273]. Therefore, the BW419Mo population comprised 3 F_2 families of 100 individuals. BW902Mo was represented by 15 F_2 families of 20 individuals.

Populations were grown under glasshouse conditions (day temperature 18^{o}C, night temperature 14^{o}C with 16 hr day length) during the autumn/winter of 2010/11. Plants were sown into 18 cm diameter pots (28/10/10), containing a compost/Intercept granular insecticide mix suitable for growing cereals. Plants were irrigated using a combination of overhead watering and, once soil in the pots had been thoroughly wetted, drip feed onto capillary matting. Subsequent to tissue being harvested for DNA extraction (2.2.5), plants underwent a prophylactic fungicide treatment with Opus-Team

(1ml/l) to protect against any possible build-up of powdery mildew infection that may bias phenotyping.

4.2.2.2 *Tissue sampling*

50—100µg young leaf tissue was harvested at GS23, snap-frozen in liquid nitrogen and stored at -80^{o}C until DNA extraction.

4.2.2.3 *Phenotyping*

Prior to harvest, plants were phenotyped for number of fertile tillers **i**.e. those tillers which had fully developed spikes which had produced grain and were fully senesced (not green). Representative spikes from each pot were photographed and assessed for row-type (two-rowed or six-rowed), double awns (presence/absence), lateral grain fill (scale from 0—5, with 0 being no lateral grain fill and 5 being full lateral grain fill), spikelets were deemed fertile when they showed evidence of grain-fill. The shape of lateral spikelets were assessed as either rounded, pointed or awned; the awned category was further sub-divided into awnletted (spikelets with awns less than 2.5cm) and fully awned, spikelets with awns greater than 2.5cm. In some cases notes were also taken on the general appearance of the spike and any environmental factors, such as irregular watering, which may have influenced the phenotypes observed.

4.2.3 ***F3 BW419Mo and BW902Mo field trial***

188 F3 individuals from the BW419Mo and BW902Mo were grown as part of the field trial previously described in 2.2.3.

4.2.3.1 *Pre-harvest phenotyping*

Pre-harvest phenotypic assessments were carried out as previously described in 2.2.3.3.

4.2.3.2 *Post-harvest phenotyping*

Individual plants within each plot of the BW419Mo and BW902Mo populations were classified according to the strength of their row-type phenotype with a numerical score being assigned to each category. Lateral spikelets were defined as fertile if there was evidence of grain production on phenotyping. Lemma awn development was assessed on a scale of fully developed, with awns greater than 2.5 cm; awnletted where the lemma awn is between 0.5 cm and 2.5 cm; pointed where no awn development is evident but the lemma tip is pointed, the least developed lemma, with no evidence of awn extension are described as rounded. Pointed, awnletted and awned phenotypes are associated with enlarged spikelets, with the rounded lemma phenotype associated with

reduced infertile spikelets. An overall spike development score was then derived for each plot from the average phenotype score.

4.2.4 ***Phenotyping of central and lateral grain***

Central and lateral grain were separated from a single representative spike of each individual from the BW902Mo and BW419Mo populations.

A MARVIN grain analyser (GTA Sensorik) was used to count and measure the grain width, length and area parameters. The Thousand grain weight (TGW) was derived from the weight of the grain in the counted sample.

Grain ratio parameters were derived by dividing the average measures of area, width and length of the lateral grain by the same measures for the central grain for each line

4.2.5 ***Genotyping***

4.2.5.1 *DNA extraction*

DNA extractions were carried out using the methodology previously described in 2.2.5.

4.2.5.2 *Row-type loci allele segregants*

At the *VRS1* locus Morex is known to contain the *vrs1*.a1 allele and Bowman *Vrs1*.b3[72] Diagnostic SNP for KASP assay design were identified by alignment of Genbank sequences AB489127.1 (*vrs1.a*.1) and EU331622.1 (*Vrs1.b*3) using Geneious6 (Biomatters Ltd.).

At the *INT-C* locus Morex has the *Int-c*.a allele and Bowman the *int-c*.b[72]. Genomic sequence data as published by Ramsay *et al.*[72] were used for the design of KASP assay.

4.2.5.3 *VRS3 KASP design*

Sequencing of *VRS3* identified a one nucleotide deletion 1774bp from the start codon in BW902 and a two nucleotide deletion 3464bp from the start codon in BW419. The sequence either side of the SNP were sent to LGC genomics for KASP assay design (Appendix 15).

4.2.5.4 *KASP Genotyping*

The KASP genotyping assays followed the protocol previously described in section 2.2.7.2.

4.2.6 ***Statistical Analyses***

The Pearson Chi-squared method was used to test associations of *VRS1* and *INT-C* with spike phenotype score in the F_3 BW419Mo and BW902Mo populations prior to the identification of *VRS3*.

F_2 Morex populations grown in the glasshouse: a linear model was fitted with *VRS1* genotype, *INT-C* genotype and cross included as fixed effects.

188 line F_3-Morex populations- for spike emergence and spike phenotype score linear models were fitted with bloc, F_2 *VRS1* genotype, F_2 *INT-C* genotype and mutant allele included as fixed effects. For the tillering analysis, the model was modified to take into account the unknown genotype of the plant in which tillering was counted; the model fitted comprised fixed effects of bloc, row-type of plant and mutant allele.

4.2.6.1 *Segregation ratio*

Genstat version 14 was used to determine the segregation ratios of *VRS1*, *INT-C* and *VRS3* using χ^2 test of association.

.

4.2.7 ***Homozygous row-type gene combination field trial***

4.2.7.1 *Design and inputs*

Germplasm:

83 entries, from the F_3 generation of the BW419Mo and BW902Mo populations representing the eight homozygous combinations of row-type genes *VRS1*, *INT-C* and *VRS3*.

Trial design and agronomy

The field trial was designed as a replicated incomplete block design using Agrobase version 35 (Agronomix). Each plot comprised two rows of plants sown over 2 m, and plots were arranged in a grid of 20 rows x 12 columns. Plots were sown using a Seedmatic seed drill on 12/04/2013, a typical applied fertiliser programme and a prophylactic fungicide treatment. 5 plants per plot were harvested into bundles at GS89[22], plants were dried with warmed air overnight to <15% moisture and stored prior to phenotyping.

4.2.7.2 *Phenotyping*

Pre-harvest

Prior to harvest plots were measured for height; 3 measurements were taken per plot, from the ground to plant collar, from the ground to the top of the spike and from the ground to the top of the awns.

At approximately GS80 plots were assessed for row-type, either two-rowed or six-rowed, and lateral spikelet shape, rounded/pointed.

Post-harvest

The fertile tillers were counted from all 5 plants harvested per plot. 1 representative main tiller spike per plant was removed for subsequent grain size analysis. The remaining tillers were threshed as a bulk sample using a Wintersteiger paddle thresher.

To separate central and lateral grain, the central grain from each harvested spike was marked with magnetic primer (Rust-Oleum) using a needle. After the primer had dried, the 5 spikes per plot were threshed together by hand and chaff discarded. Subsequently, grain were placed on a metal train, a 60mm Neodymium magnet was used to separate central grain from the lateral grain.

MARVIN grain analyser (GTA Sensorik) was used to establish grain number and size parameters on the central, lateral and bulk grain samples.

4.2.8 ***Data analysis***

Phenotypic data was analysed using REML in Genstat version 14[141].

4.2.8.1 *Central Grain, tiller and whole plant parameters*

F_2 glasshouse trial

In all cases a full interaction model of all three loci as fixed effects was initially fitted to determine significant interactions. Subsequently, non-significant effects were removed from the model prior to predicting genotype means. A list of all the models fitted can be found in Appendix 20.

Homozygous row-type combinations field trial

Fixed effects were fitted using the same methodology as the F_2 glasshouse trial. To take into account spatial and background genotype effects, the following was fitted as random effects: row.col+entry (Appendix 22).

4.2.8.2 *Lateral grain parameters*

F_2 glasshouse trial

Due to the unbalanced nature of the genotype combinations associated with the lateral grain parameters individual loci could not be fitted to the model. Instead, each genotype combination was assigned to a group, which was then fitted as a single main effect.

Homozygous row-type combinations field trial

Fixed effects were fitted using the same methodology as the lateral grain parameters from the F2 Glasshouse Trial, but a random model of row.col+entry was fitted to take into account any spatial or genetic background effects. In all cases a 5% level was used to determine significance.

4.3 *Results*

4.3.1 *Phenotypic analysis of the F2 BW419Mo and BW902Mo populations prior to the identification of VRS3*

With three row-type genes, *VRS1*, *INT-C* and *VRS3* segregating 27 different allele combinations are expected within the BW419Mo and BW902Mo F_2 populations. **Table 21** shows the differing combinations and expected ratios within a randomly segregating F_2 population.

Table 21: The expected ratios of the *VRS1*, *INT-C* and *VRS3* genotype combinations across the F_2 BW419Mo and BW902Mo populations

Genotype	Ratio
Vrs1 Vrs1,int-c.b int-c.b, Vrs3 Vrs3	1
Vrs1 Vrs1, int-c.b int-c.b, Vrs3 vrs3	2
Vrs1 Vrs1, int-c.b int-c.b, vrs3 vrs3	1
Vrs1 Vrs1, Int-c.a int-c.b, Vrs3 Vrs3	2
Vrs1 Vrs1, Int-c.a int-c.b, Vrs3 Vrs3	4
Vrs1 Vrs1, Int-c.a int-c.b, vrs3 vrs3	2
Vrs1 Vrs1, Int-c.a Int-c.a, Vrs3 Vrs3	1
Vrs1 Vrs1, Int-c.a Int-c.a, Vrs3 vrs3	2
Vrs1 Vrs1, Int-c.a Int-c.a, vrs3 vrs3	1
Vrs1 vrs1, int-c.b int-c.b, Vrs3 Vrs3	2
Vrs1 vrs1, int-c.b int-c.b, Vrs3 vrs3	4
Vrs1 vrs1, int-c.b int-c.b, vrs3 vrs3	2
Vrs1 vrs1, Int-c.a int-c.b, Vrs3 Vrs3	4
Vrs1 vrs1, Int-c.a int-c.b, Vrs3 vrs3	8
Vrs1 vrs1, Int-c.a int-c.b, vrs3 vrs3	4
Vrs1 vrs1, Int-c.a Int-c.a, Vrs3 Vrs3	2
Vrs1 vrs1, Int-c.a Int-c.a, Vrs3 vrs3	4
Vrs1 vrs1, Int-ca Int-c.a, vrs3 vrs3	2
vrs1 vrs1, int-c.b int-c.b, Vrs3 Vrs3	1
vrs1 vrs1, int-c.b int-c.b, Vrs3 vrs3	2
vrs1 vrs1, int-c.b int-c.b, vrs3 vrs3	1
vrs1 vrs1, Int-c.a int-c.b, Vrs3 Vrs3	2
vrs1 vrs1, Int-c.a int-c.b, Vrs3 vrs3	4
vrs1 vrs1, Int-c.a int-c.b, vrs3 vrs3	2
vrs1 vrs1, Int-c.a Int-c.a, Vrs3 Vrs3	1
vrs1 vrs1, Int-c.a Int-c.a, Vrs3 vrs3	2
vrs1 vrs1, Int-c.a Int-c.a, vrs3 vrs3	1

F_2 individual genotypes at the *VRS1* and *INT-C* loci were determined from the results of allele specific KASP assays. These data showed that *VRS1* showed no significant deviation from the expected 1: 2: 1 segregation ratio (χ^2 =0.14, p=0.935, 2df). In

contrast, *INT-C* showed a slight deviation from the expected 1: 2: 1 segregation ($\chi 2$=10.44, p=0.005) due to an excess of heterozygotes and a deficiency of *Int-c.a* ((6-row) homozygotes). It is not clear why the skew in *INT-C* allele frequency is observed, it is not associated with specific *VRS1* genotype combinations suggesting it is most likely a random recombination effect.

The F_2 BW419Mo and BW902Mo spikes could be divided into 5 phenotypic classes, based on their lateral spikelet development, grain-fill and awn development (Figure 47).

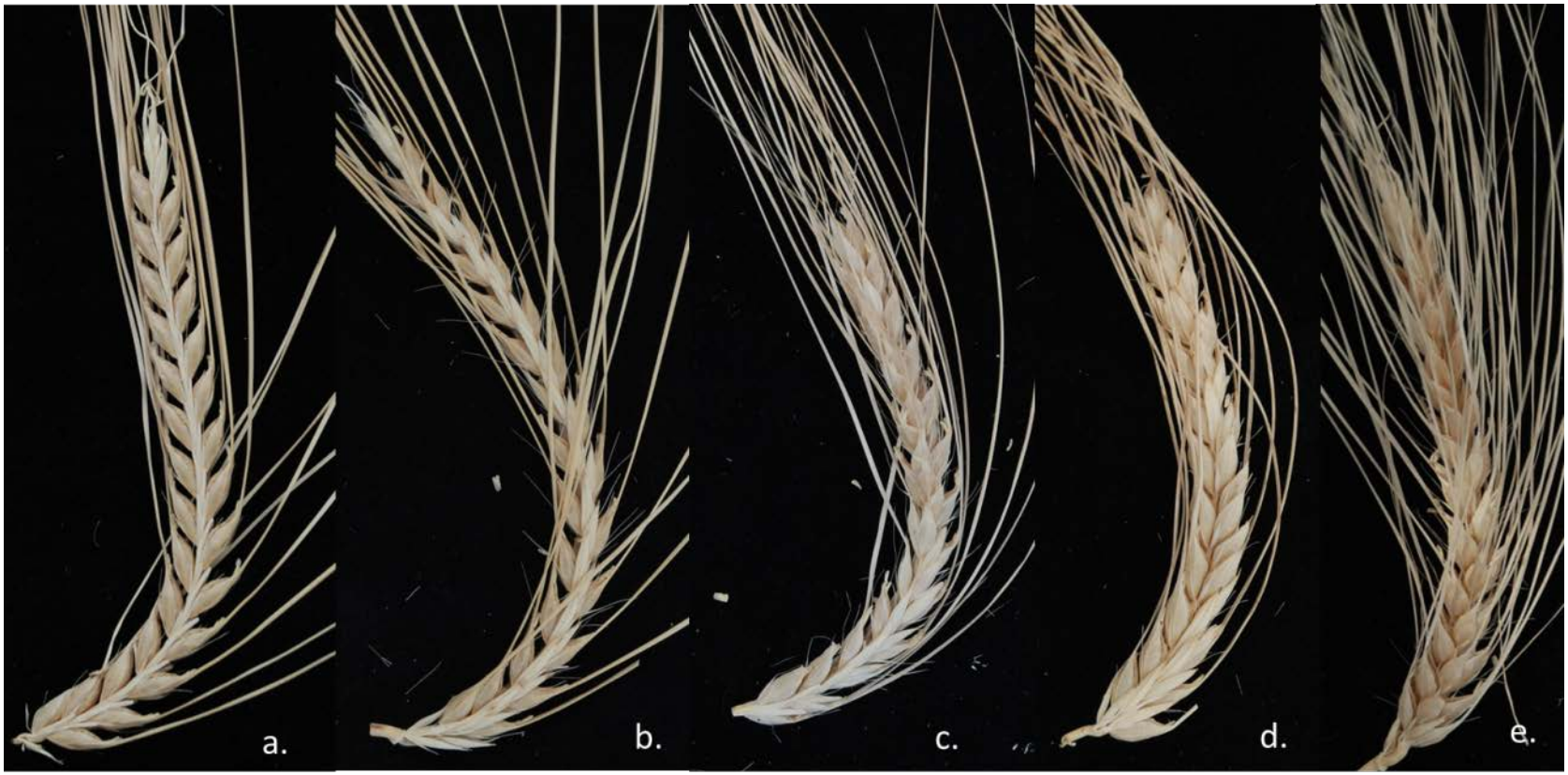


Figure 47: **The five different phenotype categories observed within the BW419Mo and BW902Mo crosses.**
a) two-rowed rounded laterals, b) two-rowed enlarged and pointed laterals, c) intermedium with partial lateral grain-fill, d) lateral grain-fill without awns , e) full six-rowed.

By combining the genotypic information at the *VRS1* and *INT-C* locus with the spike morphology phenotypes recorded, in some cases it was possible to infer the genotype at the *VRS3* locus (Table 22). However, when the *VRS1* locus was homozygous six-rowed, this was not possible. This was also the case where homozygous six-rowed *INT-C* was in combination with heterozygous *VRS1*.

Being unable to infer the genotype of *VRS3* from the F_2 phenotype meant that the F_2 population could not be used directly for linkage mapping involving *VRS3* without the non-random exclusion of individuals. Phenotyping of the individuals at the F_3 generation may provide some further evidence; however, in cases where multiple genotypes give a full six-rowed phenotype it may still not be possible to infer genotypes.

Table 22: Inference of *VRS3* genotypes from the spike phenotypes and *VRS1* and *INT-C* genotypic information collected on BW419Mo and BW902Mo F_2 individuals.

VRS1 allele [a]	*INT-C* allele [b]	Phenotype	Inferred *VRS3* genotype [c]	Total No. of individuals within Genotype Group	No. of Individuals with phenotype	Ratio
2	2	two-rowed, rounded laterals	2/H	31	22	0.71
2	2	Intermedium with pointed and awned laterals	6		9	0.29
2	6	two-rowed, enlarged pointed laterals	2/H	34	18	0.53
2	6	Intermedium with pointed and awned laterals	6		16	0.47
2	H	two-rowed, rounded laterals	2/H	76	54	0.71
2	H	Intermedium with pointed and awned laterals	6		22	0.29
6	2	six-rowed, pointed and awned laterals	unknown	43	10	0.23
6	2	six-rowed, awned laterals	unknown		32	0.74
6	2	lateral grain fill without awns	unknown		1	0.02
6	6	six-rowed, awned laterals	unknown	31	30	0.97
6	6	six-rowed, pointed and awned laterals	unknown		1	0.03
6	H	six-rowed, awned laterals	unknown	70	67	0.96
6	H	six-rowed, pointed and awned laterals	unknown		3	0.04
H	2	two-rowed, enlarged pointed laterals	2/H	61	42	0.69
H	2	six-rowed, pointed and awned laterals	6		19	0.31
H	6	six-rowed, awned laterals	unknown	45	18	0.40
H	6	lateral grain fill without awns	unknown		25	0.56
H	6	Intermedium with pointed and awned laterals	unknown		2	0.04
H	H	two-rowed, enlarged pointed laterals	2/H	170	122	0.72
H	H	six-rowed, awned laterals	unknown		19	0.11
H	H	Intermedium with pointed and awned laterals	unknown		29	0.17

[a] 2 refers to homozygous *Vrs1*.b allele, 6 homozygous *vrs1*.a allele and H heterozygous
[b] 2 refers to homozygous *int-c*.b allele, 6 homozygous *Int-c*.a allele and H heterozygous
[c] 2 refers to the *Vrs3* allele, 6 refers to homozygous *vrs3*, and H heterozygous

4.3.1.1 *F_2 BW419Mo and BW902Mo pleiotropic phenotypes*

Tiller Number

Whilst it was not yet possible to establish any effect of *VRS3* on the tiller number and lateral grain-fill phenotypes recorded in a Morex background, it was still possible to analyse the effect of *VRS1* and *INT-C.*

F_2 plants which were homozygous six-rowed at the *VRS1* locus averaged 11 tillers per plant which was significantly fewer tillers than those which were heterozygous or homozygous two-rowed at the same locus, 12 and 13 tillers respectively ($p<0.001$).

At the *INT-C* locus, no significant difference in tiller number was observed between homozygous six-rowed and heterozygous individuals; however, homozygous two-rowed individuals produced a significantly greater number of tillers compared to the other two genotypic classes, 13 tillers for two-rowed *int-c*.b homozygotes compared to 11 tillers for six-rowed *Int-c*.a homozygotes and 12 tillers for *INT-C* heterozygotes.

A significant interaction was detected between the Bowman isoline allele, *VRS1* genotype and *INT-C* genotype ($p=0.009$) (**Table 23**). Within the BW419Mo populations there was no significant difference in mean tiller number between two-rowed *VRS1* in combination with any of the three *INT-C* genotypes. In contrast, in the BW902Mo population two-rowed *VRS1* in combination with six-rowed *INT-C* produced significantly fewer tillers than *VRS1* in combination with two-rowed or heterozygous *INT-C,* suggesting perhaps an allelic interaction.

Across both BW419Mo and BW902Mo populations, there was no significant difference in the mean number of tillers from homozygous six-rowed *VRS1* in all genotypic combinations of *INT-C*; this is in contrast to reports in the literature which suggest that six-rowed *INT-C* interacts in tandem with six-rowed *VRS1* to promote tillering[72]

Table 23: BW419Mo and BW902Mo F_2 Tiller number

Due to the large differences in minimum and maximum standard error, the individual standard errors of the different transformed mean pairwise combinations are also presented (Appendix 11).

Transformed mean tiller number in the F_2 Morex populations

		INT-C Genotype			
Cross	*VRS1* Genotype	2	6	H	Mean
BW419Mo	2	3.91	3.67	3.59	3.72
	6	3.37	3.43	3.25	3.35
	H	3.66	3.38	3.62	3.55
BW902Mo	2	3.76	3.17	3.50	3.48
	6	3.23	2.89	3.34	3.15
	H	3.48	3.37	3.22	3.36
	Mean	3.57	3.32	3.42	

Untransformed mean tiller number in the F_2 Morex populations

		INT-C Genotype			
Cross	*VRS1* Genotype	2	6	H	Mean
BW419Mo	2	15.29	13.48	12.86	13.88
	6	11.36	11.76	10.54	11.22
	H	13.37	11.42	13.08	12.62
BW902Mo	2	14.15	10.07	12.22	12.15
	6	10.43	8.36	11.18	9.99
	H	12.14	11.35	10.36	11.28
	Mean	12.79	11.07	11.71	

Lateral floret fertility

Associations between genotype and lateral floret fertility could only be analysed in six-rowed spikes. Spikes were scored on a scale of 0—5 for lateral fertility (**Table 24**). Chi-squared tests of association were used to ascertain significant associations between the lateral spikelet fertility score and the genotype of individuals at the *VRS1* and *INT-C* loci (**Table 25**).

Table 24: Scoring scheme for the relative lateral spikelet fertility of the F_2 Morex populations

Lateral Spikelet Fertility	Score	Spike Phenotype Classification
None	0	Two-rowed
<10% fertile	1	Intermedium
10—50% fertile	2	Intermedium
50—90% fertile	3	Intermedium
Basal spikelets sterile	4	Six-rowed
All Fertile	5	Six-rowed

Table 25: The frequency of genotypes within each of the lateral fertility score classes. *a) INT-C, b) VRS1*

a)

		INT-C genotype			Total
		2 (*int-c*.b)	H	6 (*Int-c*.a)	
Lateral Fertility Score	1	1	0	1	2
	2	3	10	3	16
	3	24	40	25	89
	4	22	35	32	89
	5	20	55	32	107
	Total	70	140	93	303

b)

		VRS1 genotype			Total
		2 (*Vrs1*.b)	H	6 (*vrs1*.a)	
Lateral Fertility Score	1	0	2	0	2
	2	8	7	1	16
	3	27	46	16	89
	4	9	46	34	89
	5	1	13	93	107
		45	114	144	303

No evidence was found for a significant association between the *INT-C* genotype and variations in lateral fertility of the spike, (χ^2=7.75, 8 d.f. p = 0.458) with both *int-c*.b and *Int-c*.a alleles associated with a range of lateral fertility scores. This would suggest that changes in lateral fertility cannot be solely attributed to change at the *Int-c* allele.

Chi-squared tests of association found significant association between the *VRS1* genotype and the lateral floret fertility score of six-rowed spikes (χ^2=132.91, 8 d.f., p<0.001). The two-rowed allele of *VRS1* (*vrs1.b*) was associated with six-rowed spikes

with lower lateral fertility scores, suggesting that lateral spikelet fertility is possible without *vrs1* but to a lesser extent. The majority category of *Vrs1*.b (n=27) had a lateral fertility score of 3 (50—90% fertile), this most likely reflects those lines with a *vrs3* genotype, although this could only be confirmed once the gene underlying *VRS3* is identified. Six-rowed alleles of *VRS1 (vrs1.a)* were associated with high lateral fertility scores, with the predominant category being a score of 5 (n=93). The heterozygous genotype was associated with an intermediate score, with highest numbers of individuals having scores of 3 and 4 (n=46 in both cases).

4.3.2 *F_3 Morex crosses field trial of 188 individuals prior to identification of VRS3*

The primary aim of the F_3 field trial was to progeny test each F_3 family for segregation at each six-row locus to establish a genotypic ratio and thus increase precision for genetic mapping. Therefore of the 600 lines within the BW419Mo and BW902Mo F_2 populations, only the 93 lines that had been genotyped with the BeadXpress OPA were included in the F_3 field trial. However, establishing genotype from segregating phenotypes still proved difficult in the Morex crosses. The relatively small number of plants per plot, approximately 20 derived from a single F_2 spike, meant that whilst there would be a high chance of detecting at least one recessive when just one locus was segregating, the population size was only just big enough to detect a double recessive when two loci were segregating so an accurate classification, even when pooling across replications, was not always possible. Additionally, as in the F_2 plants, in homozygous six-rowed *VRS1* backgrounds, the genotypes at the *INT-C* and *VRS3* locus could not be distinguished. Table 26 identifies the phenotypes observed in the field and where possible the predicted genotype at the *VRS3* locus in the F_2 generation.

Table 26: Prediction of F_2 *VRS3* genotype from F_3 phenotype segregation in the BW419Mo and BW902 Mo populations.

VRS1 [a] F_2 allele	*INT-C*[b] F_2 allele	F_3 Phenotypes observed	Predicted F_2 *VRS3*[c] genotype
2	2	two-rowed, rounded laterals	2
2	2	two-rowed rounded laterals, Intermedium with pointed and awned laterals	H
2	2	Intermedium with pointed and awned laterals	6
2	6	two-rowed, enlarged pointed laterals	2
2	6	two-rowed, enlarged pointed laterals Intermedium with pointed and awned laterals	H
2	6	Intermedium with pointed and awned laterals six-rowed, awned laterals	6
2	H	two-rowed, rounded laterals two-rowed, pointed laterals	2
2	H	two-rowed, rounded laterals two-rowed, pointed laterals Intermedium with pointed and awned laterals six-rowed, awned laterals	H
2	H	Intermedium with pointed and awned laterals six-rowed, awned laterals	6
6	2	six-rowed, pointed and awned laterals six-rowed, awned laterals	2/H/6
6	6	six-rowed, awned laterals six-rowed, pointed and awned laterals	2/H/6
6	H	six-rowed, awned laterals six-rowed, pointed and awned laterals	2/H/6
H	2	two-rowed, enlarged pointed laterals six-rowed awned laterals two rowed, rounded laterals	2
H	2	two-rowed, enlarged pointed laterals six-rowed awned laterals two rowed, rounded laterals Intermedium, pointed and awned laterals	H
H	2	Intermedium, pointed and awned laterals six-rowed awned laterals	6
H	6	six-rowed, awned laterals	6
H	6	six-rowed, lateral grain fill without awns six-rowed, awned laterals two-rowed, rounded laterals two-rowed, enlarged pointed laterals Intermedium with pointed and awned laterals	2/H
H	H	six-rowed, awned laterals two-rowed, rounded laterals two-rowed, enlarged pointed laterals Intermedium with pointed and awned laterals six-rowed, lateral grain fill without awns	2/H/6

[a] 2 refers tohomozygous *Vrs1*.b allele, 6 homozygous *vrs1*.a allele and H heterozygous
[b] 2 refers to homozygous *int-c*.b allele, 6 homozygous *Int-c*.a allele and H heterozygous
[c] 2 refers to the *Vrs3* allele, 6 refers to homozygous *vrs3*, and H heterozygous

4.3.2.1 *Spike emergence*

Spike emergence varied across the BW419Mo and BW902Mo populations, with the earliest plots heading 70 days post sowing and the latest at 85 days. Analysis suggests that those plots which were genotyped as heterozygous for *VRS1* in the F_2 generation headed significantly earlier (1.5 days, p=0.007) in the F_3 generation than those homozygous *Vrs1* or homozygous *vrs1*. It is not clear whether this is a true effect of heterozygous *VRS1* or if the heterogeneity of the plot given the differing spike phenotypes has resulted in ambiguity in assessing plot spike-emergence.

4.3.2.2 *Tiller number*

A large proportion (7/8) of the F_3 plots would be expected to segregate for *VRS1, INT-C* or *VRS3* genotype but not necessarily for phenotype, therefore as individual F_3 plants were not genotyped, the analysis of tillering data within the Morex populations had to be limited to any association with overall spike row-type of the specific plant counted and not specific row-type loci.

As with the Barke and Bowman populations, row-type was found not to significantly affect tiller number at GS39 (P=0.293) but post-harvest tiller counts suggested a significant difference between row-types (p< 0.001) (Table 27).

Table 27: Mean tiller numbers at growth stage 39 and post-harvest for two-rowed and six-rowed in the Morex F_3 populations.
Numbers in brackets are the untransformed means, with the SED referring to differences between the transformed means.

	Two-rowed	Six-rowed	SED
GS39	2.735 (7.48)	2.686 (7.21)	0.049
Post-harvest	3.704 (13.71)	3.281 (10.76)	0.068

4.3.2.3 *Plot spike phenotype score*

Previous work has suggested that multiple six-rowed alleles of row-type loci interact to produce improved six-rowed spikes, with enhanced lateral grain-fill[132].

The plot spike phenotype score was designed to establish whether any significant observable differences could be established between the different six-rowed genotype combination of *VRS1*, *INT-C* and *VRS3*. Spikes within a plot were scored according to phenotypes detailed in Table 28. Within some plots the row-type genotypes were segregating resulting in both two-rowed and six-rowed spikes within a single plot, which is reflected in the significantly lower plot phenotype score for the heterozygous plots.

Table 28: F_3 Morex cross spike phenotyping categories and descriptions

Category	Description	Score
Full six-rowed	Regular six-rowed spike with full lateral grain fill and awns. Lateral grain appear similar size to central.	7
Weak full six-rowed	Regular six-rowed spike with full lateral grain fill and awns. Lateral grains appear smaller in size than central.	6
Lateral grain-fill, no awns	Lateral spikelets have grain-fill but no awns. Central spikelets have full grain-fill and are awned.	5
Intermediate six-rowed	Top of spike regular six-rowed with awns on central and lateral florets, awns shortening as descend spike with lower laterals pointed rather than awned, grain fill in laterals decreasing down spike with between 1 and 3 of the lower three lateral pairs infertile.	4
Weak six-rowed	Central spikelets have full grain-fill and are awned. Lateral spikelets have varying degrees of fertility and awn development. At least the lower four lateral spikelet pairs are infertile	3
Two-rowed, enlarged pointed laterals	Central spikelets are awned and fertile. Lateral spikelets are infertile, but enlarged and pointed.	2
Two-rowed, rounded laterals	Central spikelets are awned and fertile. Lateral spikelets are infertile with rounded tops.	1

The mean plot spike phenotype scores indicated that *VRS1* and *INT-C* do not account for all the variability seen in the phenotypes; the two rowed *VRS1* two-rowed *INT-C* interaction gave a mean value greater than 1 (the value assigned in phenotyping for a two-rowed rounded lateral spike). This is to be expected as the analysis does not yet take into consideration the segregation of *VRS3* within the population. Plot spike phenotype score was found to differ significantly with both *VRS1* genotype ($p<0.001$) and *INT-C* genotype ($p<0.001$), with *VRS1* having the larger effect. A significant interaction between the two was also observed ($P=0.006$). The means for the interaction are shown in Table 29.

Homozygous six-rowed *VRS1* is sufficient to produce a six-rowed spike but no significant difference was observed in the combination of *vrs1* with the three *INT-C* genotypes. Given that the *Int-c*.a allele is found in combination with *vrs1* in cultivated six-rowed genotypes, this finding suggests that *Int-c*.a is not contributing to the six-rowed morphology of the spike but a more subtle phenotypic difference, for example grain width or grain length, a refinement in grain phenotyping to quantitatively measure grain size differences will help ascertain the role of *INT-C* in six-rowed spike modification.

INT-C genotype showed a significant interaction with the allele of Bowman isoline present ($p=0.008$). Six-rowed *INT-C* in combination with BW902 gave a significantly greater mean spike phenotype score than six-rowed *INT-C* with BW419, 5.4 compared to 4.5. This would suggest the combination of six-rowed *INT-C* with BW902 gave a significantly better six-rowed phenotype, this is in line with the earlier findings of homozygous *vrs3* individuals in the F_2 BW902Bo and BW902Ba populations showing significantly enhanced lateral awn and grain development compared to those in BW419Bo and BW419Ba populations.

Table 29: The mean phenotypic spike score for the interaction between *VRS1* and *INT-C* genotypes

SED for paired means are shown below, significant differences are shaded.

VRS1 Genotype	*INT-C* Genotype Two-rowed (*int-c*.b)	Six-rowed (*Int-c*.a)	Heterozygous
Two-rowed (*Vrs1*)	2.953	3.994	2.574
Six-rowed (*vrs1*)	6.101	6.479	6.494
Heterozygous	3.461	4.351	4.168

VRS1	*INT-C*		1	2	3	4	5	6	7	8	9
2	2	1	*								
2	6	2	0.4	*							
2	H	3	0.317	0.341	*						
6	2	4	0.365	0.386	0.299	*					
6	6	5	0.43	0.449	0.376	0.417	*				
6	H	6	0.337	0.36	0.265	0.32	0.394	*			
H	2	7	0.336	0.359	0.263	0.319	0.392	0.287	*		
H	6	8	0.365	0.386	0.299	0.349	0.417	0.32	0.319	*	
H	H	9	0.293	0.319	0.205	0.274	0.356	0.236	0.233	0.274	*

4.3.3 *Row-type gene genotype segregation ratios in BW419Mo and BW902Mo F2 populations following the identification of VRS3*

The identification of the gene underlying *VRS3* (Chapter 3) enabled further genotyping of the BW419Mo and BW902Mo F_2 populations for segregation at the *VRS3* locus. This in combination with the previous genotyping at the *VRS1* and *INT-C* loci (section 4.3.1.1) partitioned the 562 F_2 individuals into 27 different genotypic combinations (Table 30).

Table 30: The observed and expected segregation ratios of the 27 genotype combinations of *VRS1*, *INT-C* and *VRS3* in the F_2 generation of the BW419Mo and BW902Mo populations.

Genotype group	*VRS1*[a]	*INT-C*[b]	*VRS3*[c]	Observed number of Individuals	Expected ratio
1	H	6	H	21	4/64
2	H	6	6	18	2/64
3	H	6	2	8	2/64
4	H	2	H	31	4/64
5	H	2	6	20	2/64
6	H	2	2	12	2/64
7	H	H	H	82	8/64
8	H	H	6	48	4/64
9	H	H	2	40	4/64
10	6	6	H	12	2/64
11	6	6	6	13	1/64
12	6	6	2	6	1/64
13	6	2	H	18	2/64
14	6	2	6	19	1/64
15	6	2	2	6	1/64
16	6	H	H	43	4/64
17	6	H	6	15	2/64
18	6	H	2	12	2/64
19	2	6	H	11	2/64
20	2	6	6	16	1/64
21	2	6	2	7	1/64
22	2	2	H	12	2/64
23	2	2	6	7	1/64
24	2	2	2	9	1/64
25	2	H	H	44	4/64
26	2	H	6	22	2/64
27	2	H	2	10	2/64
			Total	562	

[a] 6 at the *VRS1* locus refers to homozygous *vrs1*.a, 2 at the *VRS1* locus refers to homozygous *Vrs1*.b.
[b] 6 at the *INT-C* locus refers to homozygous *Int-c.a*, 2 at the *INT-C* locus refers to homozygous int-c.b.
[c] 6 at the VRS3 locus refers to homozygous mutant *vrs3*, 2 at the *VRS3* locus refers to homozygous wild-type *Vrs3*.

Chi-squared tests of goodness of fit found a significant deviation from the expected segregation ratios across the population. χ^2=59.69 (p<0.001, 26 d.f.). Further investigation found this to be due to segregation distortion at both the *INT-C* (χ^2=10.75, p=0.05, 2 d.f.) and *VRS3* (χ^2=16.56, p<0.001, 2 d.f.) loci; the reason for this is unclear as it was not attributed to any specific genotype combinations making it unlikely to be due to loci interaction; additionally, in the F2 populations used to map *VRS3* no evidence of segregation distortion was found. Segregation distortion has been reported in previous studies in barley, particularly with genes in close proximity to the centromere [274,275], as is the case for *VRS3*.

4.3.4 *Spike phenotypes of varying F_2 VRS1, INT-C and VRS3 allele combinations*

In section 4.3.1.1 the spike phenotypes observed in individuals from the F_2 generation of BW419Mo and BW902Mo populations were summarised into the following five categories:

Full six-rowed— all lateral spikelets with awns and grain-fill

Intermedium no lateral awns—lateral spikelets produced grain but no awns.

Partial lateral grainfill—grain-fill progressed up the spike, with lower lateral spikelets infertile. Awn development also progressed acropetally.

Pointed laterals— Lateral spikelets are infertile but have enlarged and pointed lemma.

Rounded laterals—Lateral spikelets are infertile with rounded lemma.

Following genotyping at the *VRS3* locus it was possible to align phenotypes with the 27 different genotypic groups within the F_2 BW419Mo and BW902Mo populations (Figure 48).

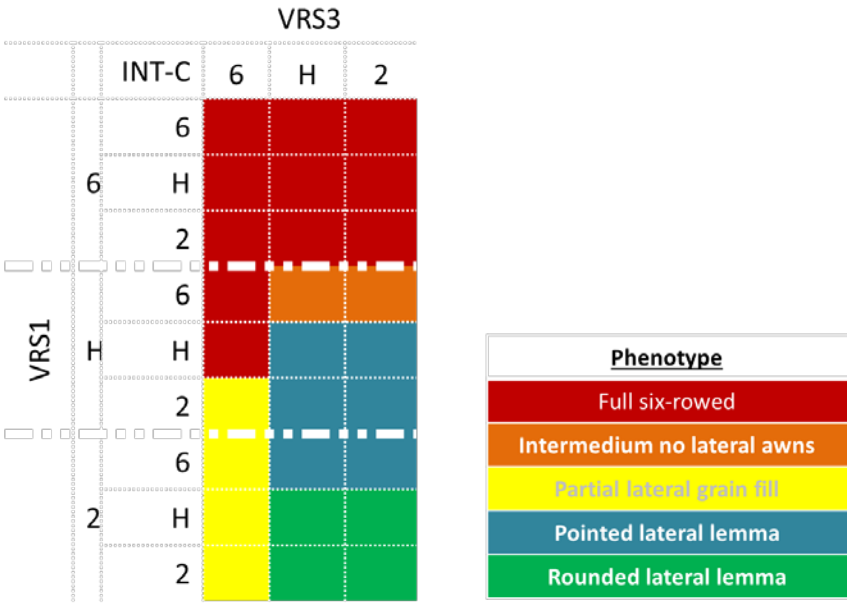


Figure 48: Association between the 27 F_2 genotype combinations of BW419Mo and BW902Mo and observed spike phenotypes.
In the case of VRS1, '6' refers to the homozygous vrs1.a genotype, '2' the homozygous Vrs1.b allele. With respect to INT-C, '6' refers to the homozygous Int-c.a allele and '2' the homozygous int-c.b allele. For the VRS3 locus '6' refers to the homozygous mutant allele of vrs3 and '2' refers to the homozygous wild-type allele of Vrs3. At all loci 'H' stands for heterozygous.

Despite only the combination of homozygous *vrs1*.a and *Int-c*.a having been identified in cultivated conventional six-rowed germplasm it is clear that there are multiple alternative genetic combinations which result in a six-rowed phenotype.

The above suggests that the formation of a six-rowed phenotype is possible in the presence of heterozygous rather than homozygous *vrs1*.a, but for this occur the *Vrs3* locus has to be in a homozygous *vrs3* state and in combination with either the heterozygous or homozygous *Int-c*.a allele. If the homozygous *int-c*.b allele is present only partial lateral spikelet grain-fill is observed suggesting the *Int-c.a* allele promotes lateral spikelet and grain development, particularly at the base of the spike. However for this promotion to occur it has to be in combination with other lateral spikelet fertility promoting loci as singularly the homozygous *Int-c*.a allele results in enlarged lateral spikelets with pointed lateral lemmas but no grain-fill.

The data also suggests that *Vrs3* is epistatic with respect to lateral awn development In combination with heterozygous *Vrs1* and *Int-c*.a, the presence of the heterozygous or wild-type *Vrs3* suppresses lateral awn development despite lateral grain-fill occurring.

This mechanism appears to be over-ridden when *Vrs1* is in its homozygous *vrs1*.a state, with fully awned lateral spikelets formed.

4.3.5 *Tillering and Grain Size analysis in F_2 BW419Mo and BW902Mo populations*

Earlier analysis of the row-type phenotypes (section 4.3.2) suggested that simple partitioning into broadly phenotypically similar groups did not provide high enough resolution to distinguish differences between genotypes with overlapping phenotypes. Therefore a more quantitative approach, utilising the MARVIN grain analyser (GTA Sensorik) to measure size parameters of both central and lateral grain separately, was used to further phenotype the F_2 spikes. These data were combined with the *VRS1*, *INT-C* and *VRS3* genotypic data to establish the effects of the three loci on the spike phenotypes.

For ease of comprehension, throughout the results alleles of *VRS1*, *VRS3* and *INT-C* will be described as either 2, H or 6. In the case of *VRS1*, 2 refers to the two-rowed allele *Vrs1*.b3 and 6 the six-rowed allele *vrs1*.a1. For the *INT-C* locus, 2 refers to the two-rowed associated allele *int-c*.b and 6 to the six-rowed associated allele *Int-c*.a. For VRS3, 2 refers to the wild-type *Vrs3* allele and 6 to the induced mutant allele *vrs3*. In all cases H refers to heterozygous. Genotypes will be written with the *VRS1* loci first, followed by the *INT-C* locus and finally the *VRS3* locus; thus the combination 262 would be a line of genotype: *Vrs1b*.3, *Int-c*.a, *Vrs3*. In cases where the change in allele in a particular genetic background is being discussed an "_" will be used to denote the locus at which the change in allele refers e.g. _2H would be the change in allele at *VRS1* in an *int-c*.b, heterozygous *Vrs3*, genetic background.

4.3.5.1 *Locus main effects on tiller and central grain parameters*

Tiller and grain size data was analysed using REML, with the simplest model of significant effects used to predict means. Initially the effect on phenotypic mean of the three loci, *VRS3*, *INT-C* and *VRS1* in isolation are described followed by any significant interactions.

<u>*VRS3* main effect</u>

With respect to tiller number the presence of 6 allele at *VRS3* led to a reduction in tiller number of -2.57 tillers compared to the 2 allele (SED=0.450) whilst the presence of the heterozygous loci produced no significant difference to the 2 allele.

The 6 allele had a negative impact on central grain number, reducing the number of central grain per spike by -3.58; in contrast the heterozygous genotype gave a marginal significant increase in central grain number (+1.15 grain) compared to homozygous 2 allele (SED=0.557).

The 6 allele also had a negative impact on central grain TGW, with a reduction of -5.24g (SED=1.015) compared to the 2 allele. The heterozygous allele was not significantly different to the 2 allele.

A similar pattern was observed for central grain area with the 6 allele reducing grain area by -0.94 mm^2 (SED=0.208). The heterozygous allele combination again was not significantly different to the 2 allele.

VRS3 was not significant as a main effect with respect to differences in the length of the central grain. However, *VRS3* did affect the central grain width with the 6 allele decreasing central grain width by 0.12 mm (SED=0.026); the heterozygous allele combination showed no significant difference to the *VRS3* 2 genotype.

It is evident that the homozygous 2 *Vrs3* genotype showed complete dominance for all central grain phenotypes and tiller number observed apart from central grain number. In this case there is some evidence to suggest the heterozygous genotype is over-dominant compared to either parental genotype but the effect is not highly significant.

***INT-C* main effect**

The 6 allele *INT-C* resulted in a -1.42 reduction in tiller number compared to the 2 allele. The heterozygous allele also resulted in a significant reduction compared to the *INT-C* 2 allele of -1.01 tillers (SED=0.466); this was however not significantly different to the 6 allele combination.

The 6 *INT-C* allele also significantly reduced central grain number, with a difference of 1.94 grains between 6 and 2 allele combinations (SED=0.548). However, unlike the tiller number phenotype, the heterozygous allele combination at *INT-C* showed no significant difference from the 2 allele combination.

Central grain TGW showed a significant reduction in the *INT-C* 6 genotype, with a difference of -5.32 g when compared to *INT-C* 2 genotype (SED=1.016). In this case the heterozygous combination gave an intermediate phenotype, with a difference of -2.6 g compared to 2 *INT-C* genotype.

The central grain area was significantly smaller in the *INT-C* 6 genotypes, with a reduction of -1.31 mm^2 compared to the 2 genotypes (SED=0.208); the heterozygous *INT-C* genotype showed no significant difference from the 2 INT-C genotype. The central grain length was not influenced by *INT-C* as a main effect but the central grain width was reduced by -0.14 mm in 6 *INT-C* in comparison to 2 *INT-C* (SED=0.0260). In this case the heterozygous *INT-C* genotype formed an intermediate phenotype, reducing grain width by -0.063 mm.

The *INT-C* locus showed variable dominance relations; with respect to tiller number the 6 allele was dominant over the 2 allele. For the central grain number and central grain area parameters the 2 allele at *INT-C* was dominant over the 6 allele and in the case of the central grain TGW and central grain width an incomplete dominance relationship was observed.

***VRS1* main effect**

The 6 allele of *VRS1* resulted in a significant reduction in tiller number of -2.29 tillers per plant compared to the 2 allele (SED= 0.4433). The heterozygous allele of *VRS1* also led to a marginal reduction in tiller number of -0.927 compared to the 2 allele.

Central grain number was significantly influenced by the allele of *VRS1*; the 6 allele gave a reduction of -4.38 central grains per spike, compared to the 2 allele (SED=0.5581). The heterozygous allele showed no significant difference to the 2 allele.

The *VRS1* homozygous 6 allele, resulted in a significant reduction in TGW, -11.13 g compared to the homozygous 2 allele (SED=1.016). The *VRS1* H combination produced an intermediate phenotype, with a TGW 2.83 g more than the 6 *VRS1* but -8.3 g less than the 2 *VRS1*.

VRS1 genotype also significantly affected central grain size parameters; comparison of 6 allele with the 2 allele found the overall central grain area was reduced by -3.57 mm^2 (SED=0.208), the central grain length by -0.355 mm (SED= 0.052) and central grain width by -0.35 mm (SED= 0.026). In all three cases the H genotype of *VRS1* gave a phenotype intermediate to the 2 and 6 genotypes; with a decrease in central grain area of -1.06 mm^2, a decrease of -0.16 mm in grain length and a decrease of -0.087 mm in central grain width.

In general the *VRS1* locus showed an incomplete dominance relationship for all phenotypes, the exception being central grain number where the 2 allele at *Vrs1* was dominant over the 6 allele.

4.3.6 ***Lateral grain parameter main effects***

As not all genotype combinations result in the production of lateral grain, it was not possible to establish the main effects of single loci on the lateral grain phenotypes. Lateral grain phenotypes were therefore analysed by comparison of the means of loci combinations and the differences discussed in section 4.3.7

4.3.7 ***Comparison of grain size and tiller phenotypic differences across the six-rowed genotype combinations of VRS1, INT-C and VRS3***

To determine if phenotypic differences were present between differing genotype combinations which resulted in a six-rowed phenotype the means of the differing genotype combinations were compared (Figure 49). A summary of the mean values can be found in Appendix 19.

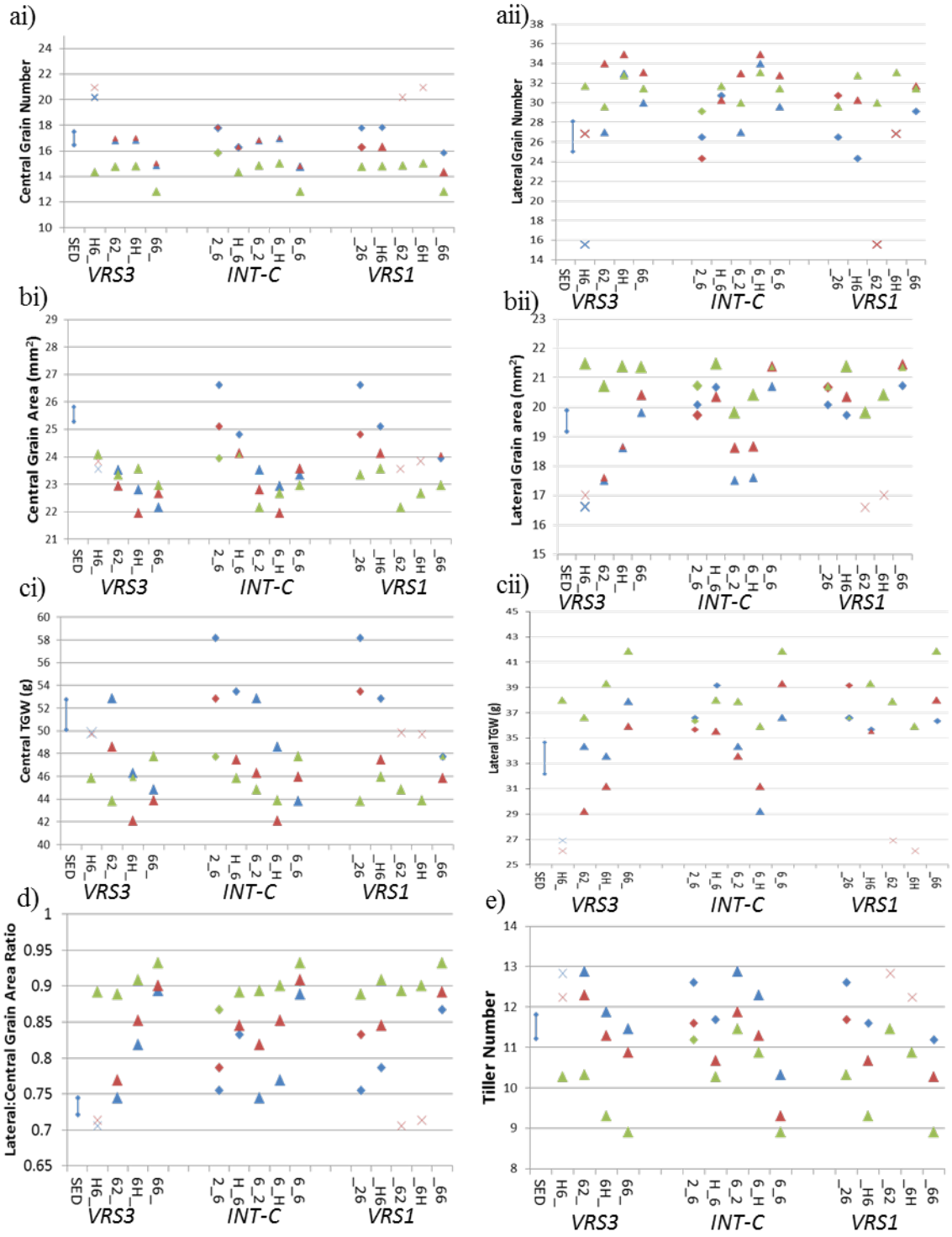


Figure 49: The change in grain parameters with change in genotype across the intermedium and six-rowed spike phenotypes.

In all cases, the x-axis represents the genetic background, with the underscore denoting the locus which is altered, ordered VRS3, INT-C and VRS1. Blue represents the 2 allele, red the H allele and green the 6 allele. Triangles represent a full six-row spike phenotype, diamonds an intermedium spike phenotype and crosses spikes which had lateral spikelets with grain-fill but no awns. Error bars represent +/- SED.

4.3.7.1 *Grain number*

The 6 allele at all three loci across all genetic backgrounds resulted in a decrease in central grain number (Figure 49 ai). The largest decreases were observed in genotype changes which resulted in a change from an awnless lateral phenotype to a full six-rowed phenotype, for example in an H6_background, the change from H6H to H66 at *Vrs3*. Within the full six-rowed phenotypes the lowest number of central grain was observed in the 666 combination.

The decrease in central grain number was coincident with an increase in lateral grain number, with changes to the 6 and H alleles at all three loci resulting in increased lateral grain number (Figure 49aii). In particular, the change from an awnless lateral phenotype to a full six-rowed phenotype resulted in a large increase in lateral grain number. As would be expected, the intermedium phenotypes generally showed fewer lateral grain than the full six-rowed phenotypes. However, the 266 intermedium allele combination resulted in a comparable number of lateral grain to the full six-rowed phenotypes, suggesting the 6 allele at *Int-c* boosts lateral grain-fill at the base of the spike.

4.3.7.2 *Grain area*

Central grain area decreased with change in phenotype from intermedium to full six-rowed (Figure 49bi). However, the central grain area between awnless lateral phenotypes and full six-rowed phenotypes was comparable. Within the full six-rowed phenotypes, the 6 allele at *VRS3* was associated with increased central grain area but the 6 allele at *INT-C* was associated with reductions in central grain area.

In comparison to the central grain area, across all genotypes the lateral grain area was smaller (Figure 49bii). Smallest lateral grains were observed in the awnless lateral grain-fill genotypes, H62 and H66. At the *VRS3* locus, the presence of the 6 allele led to significant increases in lateral grain area, particularly in the 62_ and 6H_ backgrounds. In the 66_ background, the addition of 6 at *VRS3* did significantly increase the lateral grain area over the 662 combination , average lateral grain areas for the 662 combination was 19.82 mm^2 and 21.36 mm^2 for the 666 genotype combination , a difference of +1.54 mm^2 ($p<0.001$, SED = 0.6372). Change to the 6 allele at *INT-C* was also associated with increases in lateral grain area. Interestingly, the intermedium phenotypes had comparable and in some cases significantly greater lateral grain area compared to the full six-rowed phenotypes yet maintained significantly greater central grain area.

4.3.7.3 *Thousand grain weight (TGW)*

Largest central TGW was observed in genotypes with only a single 6 allele present at any of the three loci, however of the three combinations, 622, 262, and 226, only the 622 combination resulted in a full six-rowed phenotype (Figure 49ci), with the 262 and 226 genotypes intermedium. Across the full six-rowed phenotypes, central TGW was comparable; however, a trend of increase in central TGW with change in allele at *INT-C* from 2 to 6 in combination with 6 at *VRS1* was detected.

Lateral TGW was lower than the central TGW (Figure 49cii); this difference was at its greatest in the awnless lateral phenotypes with the lateral TGW approximately half the weight of the central TGW. In general the presence of the 6 allele at all loci was associated with an increase in lateral TGW; this was particularly pronounced in the 666 genotype combination which showed the maximum lateral TGW.

4.3.7.4 *Lateral:central grain area ratio*

Across the genotype combinations, the lateral:central grain area ratio varied from 0.7 to 0.95 with the lowest grain area ratios associated with the awnless lateral phenotype (Figure 49d). At all loci the change to the 6 allele resulted in an improved grain area ratio. The change from the current cultivated six-rowed model of 662 to 666 did increase the central to lateral grain ratio but this difference was not significant.

4.3.7.5 *Tiller number*

With increased numbers of 6 alleles present in a genotype combination, the tiller number per plant decreased, with lowest tiller numbers observed in the 666 genotype combination (Figure 49e). In particular, the presence of the 6 allele at *VRS3* in varying combinations of *INT-C* and *VRS1* resulted in a larger decrease in tiller number compared to the 2 or H allele at *VRS3*.

4.3.8 ***Homozygous combinations trial***

Whilst the F_2 genotypes give a good insight into the effects of *VRS1*, *INT-C* and *VRS3* on lateral grain parameters, the population was grown under glasshouse conditions. To observe if the same effects were evident under field conditions, accessions representing the eight *VRS1*, *INT-C* and *VRS3* homozygous gene combinations were grown in a replicated field trial.

4.3.8.1 *Representative Genotype combinations*

Table 31: The number of F_3 lines which represented each of the treble homozygous genotype combinations in the field trial

Genotype	Number of Lines
222	9
226	6
262	7
266	14
622	5
626	18
662	6
666	13

4.3.8.2 *Central and whole plant grain parameters*

In addition to the central and lateral grain size parameters measured on 5 main tillers per plot, the same parameters were measured on a sample of grain from a bulk 5 plant grain sample to give an idea of within plant variation.

***VRS3* main effects**

The 6 allele at *VRS3* resulted in a significant decrease with respect to tiller number, reducing tiller number by -1.5 tillers compared to the 2 allele (SED=0.6393).

Presence of the 6 *VRS3* allele significantly reduced central grain number by -2.06 compared to the 2 allele (SED=0.4369). This was also the case for the central grain area and width parameters, with both significantly reduced by -0.88 mm^2 (SED=0.2991) and -0.058 mm (SED=0.01974) respectively. *VRS3* as a main effect did not significantly affect central grain length, however it did form a significant interaction with *VRS1*, which will be discussed later. The changes in grain area translated into a significant reduction in TGW, with the 6 allele -3.37g less than the 2 allele (SED=0.6320).

With respect to whole plant parameters, the 6 allele at *VRS3* reduced grain area by -2.26 mm^2 (SED= 0.3172), grain length by -0.354 (SED=0.1193) and grain width by -0.165 mm (SED=0.02066). The TGW was also reduced by -6.87 g (SED=0.6731).

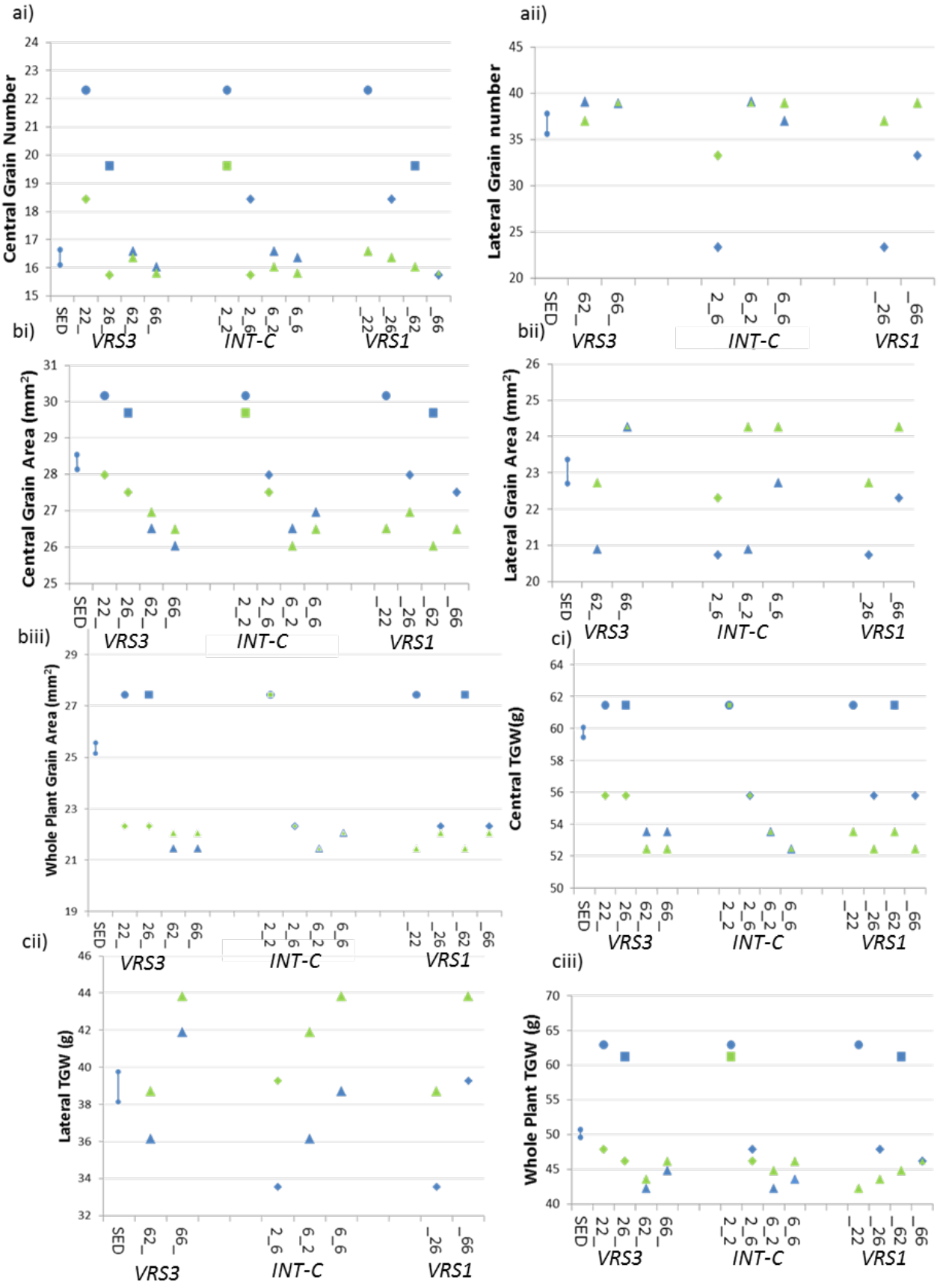
INT-C main effects

Tiller number was significantly reduced in the presence of 6 *INT-C*, with a reduction of -1.21 tillers compared to 2 *INT-C* (SED=0.6401). This was also the case for Central grain number, with a reduction of -1.62 grain (SED=0.4136). The allele at *INT-C* was not significant as a main effect across all the central grain size parameters, area, length and width, and the central TGW. The same was true of the whole plant grain size and TGW phenotypes.

VRS1 main effects

A reduction of -3.2 tillers per plant was attributed to the change between the 2 and 6 allele of *VRS1* (SED=0.6519) and the number of central grains per spike were reduced by -2.82 grain (SED=0.4271). The change from 2 to 6 at *VRS1* also resulted in a reduction in central grain area of -2.28 mm^2 (SED=0.2911), central grain length of -0.368 mm (SED=0.1076) and central grain width of-0.162 mm (SED=0.02015). The central TGW was reduced by -5.64 g (SED=0.6320).

The change from 2 to 6 at *VRS1* resulted in a significant reduction across all whole plant parameters; grain area: -3.12 mm^2 (SED = 0.3172); grain length: -0.422 mm (SED = 0.1193); grain width: -0.305 mm (SED = 0.02018) and TGW: -10.37 g (SED = 0.6575).


ai)
Central Grain Number
SED
22_
26_
62_
66_
2_2
2_6
6_2
6_6
_22
_26
_62
_66
VRS3
INT-C
VRS1
aii)
Lateral Grain number
bi)
Central Grain Area (mm²)
bii)
Lateral Grain Area (mm²)
biii)
Whole Plant Grain Area (mm²)
ci)
Central TGW(g)
cii)
Lateral TGW (g)
ciii)
Whole Plant TGW (g)

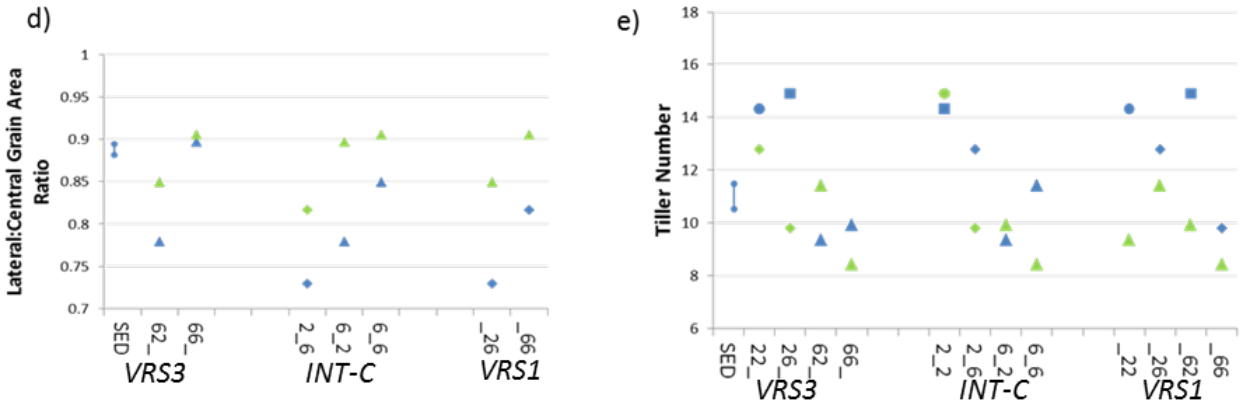


Figure 50: Comparison of the change in allele with the mean phenotype across the homozygous combinations of *VRS3, INT-C* and *VRS1*.

In all cases the x-axis represents the genetic background, with the _ denoting the allele being changed. Blue symbols represent the 2 allele and green symbols the 6 allele. Circles represent two-rowed spikes with rounded laterals; squares: two-rowed spikes with pointed laterals; diamonds intermedium spikes and triangles full six-rowed spikes. Error bars represent +/- SED.

A summary of the mean values for each phenotype/genotype combination can be found in Appendix 21.

4.3.8.3 *Spike emergence and spike length*

No significant difference was observed between lines with respect to spike emergence or spike length (data not shown).

4.3.8.4 *Grain number*

The change from two-rowed rounded laterals to two-rowed pointed laterals resulted in a significant decrease in central grain number (Figure 50ai). A further decrease was observed in the change from two-rowed pointed to full six-rowed spikes. The central grain number of the varying genotype combinations which resulted in a full six-rowed phenotype were comparable; the data suggested the 6 allele at *VRS3* and the 6 allele at *INT-C* resulted in fewer central grain compared to the 2 allele but this reduction is not significant. Change to the 6 allele at *VRS1* was associated with significant reductions in central grain number across the phenotypes except in _66 genotype combination where no significant difference was detected with change in allele at *VRS1*.

Across the full six-rowed spike phenotypes, change in allele at *INT-C* or *VRS3* did not result in a significant change in lateral grain number (Figure 50aii). However, within the intermedium phenotypes, the addition of 6 *INT-C* to the 2_6 genotype background resulted in a significant increase in lateral grain number, consistent with the 6 *INT-C* allele promoting lateral grain-fill at the base of the spike. The subsequent addition of 6

VRS1 to this genotype combination resulted in a further increase in lateral grain number and change to a full six-rowed spike phenotype.

4.3.8.5 *Grain area*

Full six-rowed spikes were associated with smaller central grain areas compared to both two-rowed and intermedium spike phenotypes, with the largest differences observed between the two-rowed and full six-rowed phenotypes (Figure 50bi).

In the presence of the 2 allele at *VRS1* the 6 allele at *VRS3* significantly reduced central grain area. This trend was flipped in the presence of the 6 allele at *VRS1*, with the 6 allele at *VRS3* showing increased grain area, however the increase was not significant over the 2 allele. In all backgrounds, the 6 allele at *INT-C* was associated with a slight decrease in central grain area compared to the 2 allele.

Lateral grain area was smaller than the central grain area, with a maximum of approximately 24 mm^2 compared to 27 mm^2 in the largest central grain of full six-rowed spikes (Figure 50bii). The change to the 6 allele at either *INT-C* or *VRS1* promoted lateral grain area across all genetic backgrounds. The effect of change at *VRS3* with change in lateral grain area was only observed in the 62_ background. The 662 and 666 genotypes showed no significant difference in lateral grain area.

As would be expected, whole plant grain area was intermediate to central and lateral grain area (Figure 50biii). Whole plant grain area followed a similar pattern to the central grain area with the 6 allele at *VRS3* decreasing grain area in a 2 *VRS1* background but increasing grain area in 6 *VRS1* background. Change in allele at *INT-C* did not affect overall whole plant grain area. In cases where the change from 2 to 6 at *VRS1* resulted in a change of spike phenotype from two-rowed to full six-rowed, there was a significant decrease in whole plant grain area. However, in cases where the change in spike phenotype was from intermedium to full six-rowed, no difference in whole plant grain area was observed.

4.3.8.6 *TGW*

Central TGW decreased significantly with change from two-rowed to either intermedium or full six-rowed spikes. (Figure 50ci). Across all genotype combinations, the 6 allele at *VRS3* resulted in a decrease in central TGW compared to the 2 allele. However, in the case of the full six-rowed spike phenotypes this difference was insignificant. Change of allele at *INT-C* did not affect central TGW.

In contrast to central TGW, lateral TGW was increased significantly with the change from 2 to 6 at the *INT-C* allele. An increase was also observed with change at the *VRS3* locus, but this increase was insignificant (Figure 50cii).

The change in allele at the *INT-C* locus also contributed to a significant increase in whole plant TGW in the full six-rowed spike phenotypes. Whilst there was an increase in whole plant TGW in the full six-rowed spike phenotypes associated with change to the 6 allele at *VRS3*, this increase was not significant (Figure 50ciii).

4.3.8.7 *Area ratio*

An improved lateral:central grain area ratio was associated with change to the 6 allele at all three loci. In the case of *VRS3*, the change was significant in the 62_ background but not the 66_ background. The 6 allele at *INT-C* was associated with a significant enhancement in lateral:central grain area ratio across all three genetic backgrounds; the same was true of the 6 allele at the *VRS1* locus (Figure 50d).

4.3.8.8 *Tiller number*

Tiller number reduced with increased lateral grain fertility, with the greatest number of tillers observed in the two-rowed genotype combination and the fewest in the 666 genotype combination. Across the full six-rowed phenotypes, no significant difference in tiller number was observed with change in allele (Figure 50e).

4.3.9 ***Summary of phenotypic changes associated with change from the current cultivated 662 six-rowed model to the alternative 666 model***

Comparison of 662 and 666 in both glasshouse and field environments has found that in both environments the mean lateral grain area was greater in the 666 genotype combination compared to 662; although the difference was only significant in the more stressed glasshouse environment. To establish the changes in grain parameter which underlied the change in lateral grain area the distribution of the varying grain width and grain length size fractions from 662 and 666 spikes was established (Figure 51).

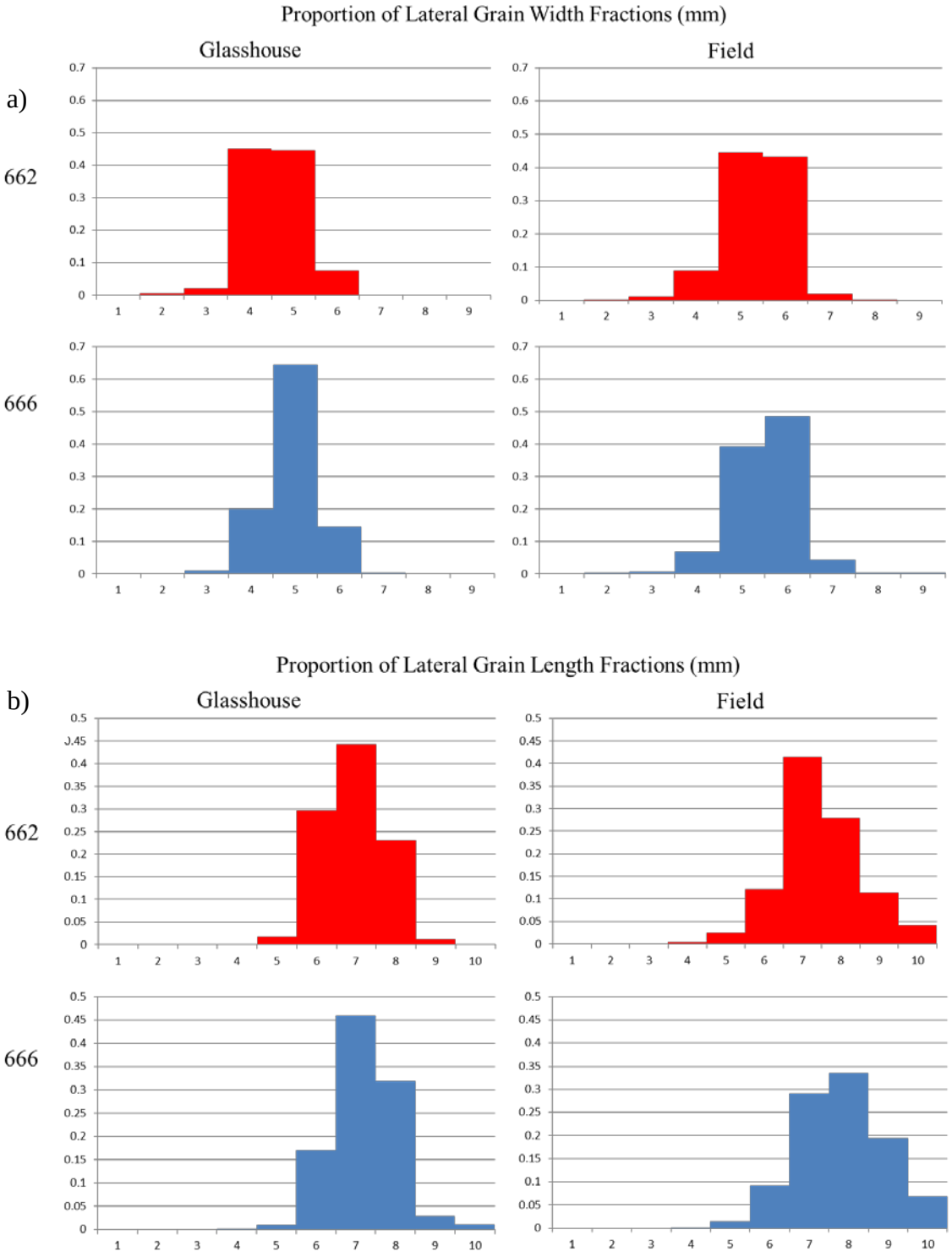


Figure 51: Comparison of the distribution of lateral grain size a) width and b) length fractions of the genotypes 662 and 666.
Red bars represent the 662 genotype and blue the 666 genotype.

Category	Width Fraction (mm)	Length Fraction (mm)
1	1.50 >≤ 2.00	≤ 3.00
2	2.00 >≤ 2.50	3.00 >≤ 4.00
3	2.50 >≤ 3.00	4.00 >≤ 5.00
4	3.00 >≤ 3.50	5.00 >≤ 6.00
5	3.50 >≤ 4.00	6.00 >≤ 7.00
6	4.00 >≤ 4.50	7.00 >≤ 8.00
7	4.50 >≤ 5.00	8.00 >≤ 9.00
8	5.00 >≤ 5.50	9.00 >≤ 10.00
9	5.50 >≤ 6.00	10.00 >≤ 11.00
10		11.00 >≤ 12.00

Lateral grain width

With respect to lateral grain width, in the glasshouse environment, fractions 4 and 5 (3.00-3.50 mm and 3.50-4.00 mm) were the dominant fractions, present in approximately equal proportions. However, in comparison in the 666 genotype, fraction 5, (3.50-4.00 mm) predominated with 62% of the sample, in this case fraction 4, corresponded to 20% of the sample. The change to the predominance of one grain fraction in the 666 genotype suggests addition of the 6 allele at VRS3 does result in an increase in uniformity of the grain sample under glasshouse conditions. Under field conditions the change from 662 to 666 resulted in an increase in grain within fraction 6 and 7, and a slight decrease in fractions 4 and 5, but the change is not as pronounced as the glasshouse environment.

Lateral grain length

The change in genotype from 662 to 666 did affect lateral grain length distribution but the shift was not as pronounced as the shift associated with lateral grain width. In the glasshouse environment, both genotypes grain length was predominated by fraction 7, 8,00 - 9.00 mm. There is evidence for a slight increase in lateral grain length with change from 662 to 666 with a reduction in the proportion of fraction 6 and an increase in fraction 8. In the field environment, again fraction 7 predominated in the 662 genotype,; in the 666 genotype, fraction 7 made up a smaller proportion of the grain, with fractions 8, 9 and 10 showing increased proportions, suggesting a more variable grain sample with respect to lateral grain length in the 666 genotype compared to 662.

Investigation of the grain size distributions supports the hypothesis that addition of the 6 allele at *VRS3* to the current cultivated six-rowed model of 662 does improve lateral grain width and grain length; in particular under stressed environmental conditions, the uniformity of lateral grain width is improved as illustrated by the predominance of a single grain width fraction.

4.4 *Discussion*

4.4.1 ***Phenotypic assessment prior to the identification of Vrs3***

Prior to the identification of the gene underlying *VRS3* through genetic mapping (Chapter 3), the prediction of genotype through the analysis of phenotype proved challenging in the three-way row-type gene segregating populations, particularly in lines with homozygous *vrs1*.a due to overlapping phenotype and genotype combinations.

Despite the challenges in identifying genotypes it was still possible to realise some conclusions as to the effect of *VRS1* and *INT-C* on the phenotypes assessed.

The F_2 Morex populations suggested both *VRS1* and *INT-C* loci significantly affected tiller number, as in both cases, the six-rowed allele was associated with a significant reduction in tillers.

The *INT-C* locus of barley has previously been associated with differences in tillering, in comparison of the Bowman isoline of mutant *int-c*.5 with Bowman; the *int-c*.5 mutant has previously been shown to result in increased tillering at GS24—25[72]. This finding is in contrast with the results from the *VRS3*/Morex populations suggesting that the six-row associated genotype *Int-c*.a represses tillering in comparison to wild-type *int-c*.b. The tiller assessments were made at a later of stage of plant development in the Vrs3/Morex population compared to the earlier study of *int-c*.5. The difference between the two studies could suggest six-rowed *Int-c* promotes increased tillering initially but any differences are lost due to differences in rates of tiller mortality associated with the six-rowed phenotype.

Spike fertility scores

The lateral fertility scores in the F_2 Morex populations, suggest *VRS1* is significantly associated with increased lateral fertility but *INT-C* is not..

Analysis of the phenotypes prior to the identification of *Vrs3* suggested that change from the *Vrs1*.b to *vrs1*.a allele at the *VRS1* locus promotes lateral spikelet fertility consistent with the findings of Komatsuda *et al.* that mutation at the *Vrs1* locus is sufficient to cause a change from two-rowed to six-rowed morphology[78]. The six-rowed *vrs1*.a allele was associated with a range of lateral fertility scores suggesting that other factors, most likely *INT-C* and *VRS3* are altering the level of lateral spikelet fertility

observed. The lack of significant association between *INT-C* and changes in lateral spikelet fertility is surprising given the consistent selection of *int-c*.a with *vrs1*.a within cultivated six-rowed germplasm[72] , however it is consistent with previous reports which suggest that the cultivated six-rowed allele *int-c*.a is only able to produce a true six-rowed phenotype when in combination with other enhancers of lateral floret fertility[88,266] .

4.4.1.1 *Post Vrs3 identification phenotypic analysis*

The identification of *Vrs3* and subsequent genotyping for the specific allele present in the BW419Mo and BW902Mo populations enabled definitive genotypic characterisation at the three row-type loci and subsequent division and further detailed characterisation of phenotypes.

In general the addition of the 6 allele at a row-type locus proved advantageous with respect to the lateral grain parameters but not with respect to the central grain parameters. In fact with the increase in size of lateral grain, there was in most cases an associated reduction in size of the central grain which would produce a more uniform sample, provided that screenings do not increase.

Of particular note was the significant reduction in tiller number with the increase in number of loci present in a 6 allelic state perhaps reflecting a balance being sought by the plant with respect to storage capacity and availability of nutrients for starch synthesis. In the F_2 glasshouse plants there was a large decrease in tiller number across all alleles of *INT-C* and *VRS1* when *VRS3* was present as a 6 allele, which would suggest that *Vrs3* promotes tiller outgrowth.

4.4.2 ***The current cultivated six-rowed model***

Previously it has been suggested that the reason behind the recurrent selection of the *Int-c*.a allele with the *vrs1*.a allele in the breeding of six-rowed cultivars was due to the *Int-c*.a allele promoting production of an increased number of tillers per plant compared to the *vrs1*.a allele alone[72]. The results here suggest that rather than increasing the number of tillers the presence of the *Int-c*.a allele with *vrs1*.a further reduced tiller number in the glasshouse conditions and caused no significant difference to tiller number in field conditions.

The *Int-c*.a allele had also been attributed to increased fertility and grain size in the lateral spikelets[84,88,276]. The results here support this, as under both field and glasshouse

conditions the *Int-c*.a allele increased the lateral grain width and lateral grain area, resulting in a more uniform grain size across the spike. Under field conditions this translated to a significant improvement in the whole plant TGW suggesting improvement across all tillers and not just the main spikes.

4.4.3 ***Improved cultivated six-rowed model***

In their earlier work, Lundqvist *et al.* suggested the formation of super 'six-rowed' spikes by the pyramiding of alleles at independent intermedium loci [134]. The results presented here do support this hypothesis, with the 662 and 626 combinations producing in general larger lateral grain parameters than the 622 combination. However, the addition of the 6 allele at *VRS3* to the 66_ background to form the 666 genotype, produced only significant increases in lateral grain width and lateral grain area under glasshouse conditions. Under field conditions the same change resulted in increases but these were not significant compared to the current cultivated six-rowed model, the 662 allele combination. Under the glasshouse conditions, the distribution of grain width was reduced in the 666 genotype compared to the 662 genotype, indicative of a more uniform lateral grain sample.. The grain produced in the field were larger overall than those produced by the greenhouse grown plants which would suggest the conditions in the field were promoting plant growth and grain production to a greater extent than the glasshouse. This would suggest that the 666 model could be a better model for six-rowed barley under stressed conditions.

4.4.4 ***Genetic control of grain size***

In barley, little is understood about the genetic control of grain size and shape; 5 induced mutant globosum loci (*glo-a (*4H), *glo-b* (5HL), *glo-c* (2H), *glo-e* (1HL) and *glo-f* (5HL)) have been identified which result in either reduction or increase in grain size with rounding of the kernel in many cases, however, none of these have yet been identified to the gene level[277]. Additionally, QTL for grain size and shape have been detected in a number of crosses; in particular in a double-haploid population from the cross between the six-rowed cultivar Morex and two-rowed cultivar Harrington, Ayoub *et al.* identified QTL associated with grain size and shape in the region of the *VRS1* and *INT-C* loci[271]. This supports the findings here that allele changes at the *INT-C* and *VRS1* loci do result in changes in grain size. Their study also identified a QTL on chromosome 1H but this was distal to the region of *VRS3*.

In rice, *An-1* and *An-2*, a basic-helix-loop-helix (*bHLH*) transcription factor[278] and *Lonely-guy like*[279] (*LOGL*) gene respectively have both been associated with the regulation of awn development, grain size and grain number. The *An-1* locus is associated with awn initiation and *An-2* awn extension[279]. This study linked the expression of *An-2* with increased levels of cytokinin, suggesting this to mediate the awn development process; this was also linked to reduced tiller number[279]. The initiation of awn development and increased grain size associated with *An-1* was linked to increased cell differentiation[278]. However, the increased expression of *An-1* was associated with reduced meristem activity, resulting in fewer grain per panicle[278]. Whilst functional orthologues of *An-1* and *An-2* are yet to be identified in barley, these studies provide potential clues as to mechanisms underlying the differences in awn development, grain size and number observed in this study.

It is not clear whether the increases in grain size parameters observed are due to the increased size of the sink for starch storage, i.e. increases in the lemma and palea, or increases in rates of grain-fill. With respect to *vrs3*, the mutant phenotypes suggest altered morphology of the lemma and palea in both rice and barley, which would suggest an increase in sink size is more likely. This would also be supported by the enlargement of lateral spikelets by the *Int-c*.a allele in an otherwise two-rowed background. Cottrell *et al.* suggested that the relative size of the spikelets is set even as early as the double ridge stage of plant development[280]. In future studies it would therefore be interesting to isolate the spikelets prior to anthesis to determine whether the differences are due to grain-fill or sink size. This would also assist in the determination of the cause in reduction of size in the central spikelets —is it due to physical space constraints within the spike at grain-fill or is the decision made earlier in plant development?

The potential interaction between *VRS3* and *INT-C* is an interesting finding of this work. With regard to the phenotype, the addition of the 6 allele at *INT-C* to the 6 *vrs3* (266) locus resulted in the lateral spikelets at the base of the spike to produce grain where in a typical 226 *vrs3* genotype the spikelets would have been reduced and infertile. Additionally, in the glasshouse environment the increase in lateral grain size associated with the 6 allele at *VRS3* appeared to be correlated with the genotype at the *INT-C* locus, particularly with respect to lateral TGW and lateral grain width, with the greatest improvements observed with *INT-C* in a 6 allele state.

The maize orthologue of *INT-C* is *TEOSINTE BRANCHED 1*; this gene has been associated with the domestication of maize from its wild ancestor teosinte. Despite no polymorphism existing within the coding region of the gene, polymorphisms in an upstream regulator result in two-fold higher expression in maize compared to teosinte [126]. In maize the expression of *Tb1* results in the suppression of lateral bud outgrowth, resulting in single tillered plants, compared to the highly branched teosinte.

Within the teosinte spike, two types of spikelet are formed, sessile and pedicellate, with only the sessile spikelets developing to form kernels; investigation of the expression pattern of *tb1* found it only to be present at the base of the repressed pedicellate spikelets suggesting a role in repressing their outgrowth[281]. This could potentially explain the difference in phenotype observed between the lateral spikelets of the *Int-c*.a and *int-c*.b spikelets in a two-rowed background, with the lateral spikelet development suppressed by *int-c*.b in the rounded laterals but de-repressed in the enlarged pointed lateral spikelets of *Int-c*.a plants. The cell specific expression pattern of *INT-C* is yet to be determined in barley but could provide some evidence to support or discount this hypothesis.

The Arabidopsis orthologue of *TB1, TBL1* has been demonstrated to show differential levels of expression across different buds within the plant; buds lower in the plant showed levels four-times higher than those further up[282]. Could this differential expression pattern explain the difference in phenotype observed in lateral spikelet development across the *vrs3* spike, with the mutant *vrs3* able to overcome the potentially lesser repressive nature of *INT-C* in the upper lateral spikelet meristems but not in the lower. Further work investigating the tissue specific expression patterns of both *INT-C* and *VRS3* would be required to determine this, as would tissue specific qRTPCR studies to determine differential expression levels.

In both teosinte and sorghum the levels of *tb1* expression have been shown to differ with respect to environment, particularly plant density. In higher density plant stands, the amount of shade is increased and therefore the ratio of red:far red light transmitted to the plant is also reduced. In these conditions teosinte has been observed to show reduced tillering in correspondence with increased levels of *tb1*[283]. A similar result was observed in Sorghum, with lateral bud outgrowth suppressed and *tb1* levels increased in response to increased far-red light[284]. Additionally they noted the same response in *phytochrome-b* mutant plants, suggesting the response of *tb1* is mediated through the

phytochrome-b shade response receptor[285]. Woodward reported in his studies of the *Int-C* locus that the increased fertility, I^h allele, was environmentally sensitive with increased lateral grain-fill observed in plants planted at low plant density[88]. Whilst it is not yet known how the I^h allele relates to the previously characterised alleles of *Int-c* it could provide a possible explanation for the differences in the influence of the *INT-C* locus observed in the glasshouse compared to the field. In the glasshouse the light levels were lower and plant density higher and therefore if true it would be expected that the *INT-C* levels of expression would be higher. Further evidence for the Phytochrome-b pathway influencing spikelet development is found at the *GS5* , grain size locus in rice. High levels of expression of the serine carboxypeptidase-like protein (*SCPL*) at this locus have been associated with the increase in volume of lemma and palea[286]. However, *GS5* expression was found to be significantly suppressed in a *phytochrome-b* mutant background[286].

For a full six-rowed phenotype to be formed *VRS1* has to be at least in a heterozygous state, and in combination with either six-rowed *INT-C* or six-rowed *VRS3*; it is not clear whether this is because they are part of independent regulatory networks suppressing different parts of the lateral spikelet development or regulated by each other. There is some evidence from teosinte and Sorghum to suggest that *INT-C* could be controlling VRS1 expression. *GRASSY TILLERS 1 (GT1)*, an HD-ZIP1 transcription factor paralogous to *VRS1* which promotes apical dominance and suppresses bud outgrowth was found to be suppressed in a *tb1* background[287]. However, in the same study, Whipple *et al.* noted that the expression of *tb1* in a *gt-1* background was not changed yet in the absence of *GT-1*, *tb1* expression was not sufficient to suppress the outgrowth of the lateral buds[287], much the same as what is observed in the formation of six-rowed barley in the 622, *VRS1*, *INT-C,VRS3* gene combination. During the course of this experiment, the gene underlying a fourth row-type locus, *VRS4*, was identified as a LOB domain containing transcription factor orthologous to the maize *Ramosa2* gene. The mutant form of *vrs4* results in multi-floretted phenotypes due to loss of spikelet determinacy and also a down-regulation of *Vrs1* suggesting that *Vrs4* also controls the expression of *Vrs1*[130]. How or if *VRS3* fits into the pathway still remains unclear.

A number of early studies describe a *Hordeum* intermedium type of barley, having lateral grain-fill but no awns[84,88]; there was some debate as to whether this was a fixed genotype or a hybrid[84,87,88]. The phenotype—genotype combinations observed in this

study support the findings of Leonard[87] that the intermedium phenotype is a hybrid, heterozygous at the *VRS1* locus and homozygous for the six-rowed allele of *INT-C*.

Whilst this study has shown under some conditions the inclusion of a 6 allele at *VRS3* does result in an improved six-rowed phenotype, it has only considered a relatively small proportion of the available alleles. In terms of *VRS1*, within European six-rowed cultivated germplasm both *vrs1*.a2 and *vrs1*.a3 alleles are more prevalent, therefore it would be beneficial to observe if the inclusion of these alleles in place of *vrs1*.a offer any further improvement. At the *INT-C* locus, currently only a single *Int-c*.a allele has been identified, however inclusion of 'stronger' six-rowed induced mutant alleles, e.g. *int-c*.5, within the three way combination to see if they promote lateral grain enhancement further would be of interest. The BW419 and BW902 mutants of *vrs3* are the phenotypically strongest, both are frameshift mutations and therefore likely to result in the production of non-functional proteins. It would therefore be interesting to include some of the non-synonymous mutants to observe what effects they may have in conjunction with *INT-C* and *VRS1*.

Due to the nature of the crosses available this study has investigated the effects in a spring six-rowed background. In terms of commercial plant breeding, in Europe the six-rowed spring market is very small, largely restricted to Scandinavia[256]; in a separate project outwith this thesis the investigation of the effect of the addition of the *vrs3* allele to cultivated six-rowed winter germplasm has begun to be investigated, the preliminary results of which suggest that a multi-floretted rather than regular six-rowed spike is formed.

5. Row-Type Loci Gene Expression

5.1 *Introduction*

The gene underlying *VRS3* has been successfully identified as a JmJC Histone demethylase (Chapter 3); however, in barley nothing is yet known about its pattern of expression. A clearer understanding of the expression of *Vrs3*, particularly during inflorescence development, may help ascertain its demethylase targets and the genetic network within which *Vrs3* functions to produce the row-type phenotype.

A number of different techniques have been previously used to successfully determine the expression patterns of loci; these include analysis of expression of single loci using quantitative Reverse Transcription PCR (qRTPCR) and more genome-wide analyses such as micro-arrays and RNA-seq.

qRTPCR targets the amplification of small amplicons (approx. 100bp) of the gene of interest from cDNA samples. The two principal methods used in qRTPCR are either SYBR green or hydrolysis probes. In the SYBR green method, during gene specific PCR, SYBR green dye binds synthesised double stranded DNA and emits fluorescence[288]. The level of fluorescence emitted is proportional to the amount of double stranded DNA synthesised. The number of PCR cycles taken to reach a threshold fluorescence is proportional to the amount of cDNA of the gene of interest in the starting reaction. In the hydrolysis probe method, in addition to the gene specific primers, a third hydrolysis probe binds to the amplicon; this probe is degraded during amplification, with the subsequent release of fluorescence detected and quantified. The number of cycles required to reach a threshold of fluorescence is used to determine the relative levels of expression[289].

Micro-arrays are comprised of thousands of oligonucleotide probes attached to a solid support. The probe sequences are unique sequences which represent expressed portions of individual genes. Fluorescently labelled samples of cDNA are applied to the array with the amount of fluorescence detected at a probe proportional to the level of gene-expression[290]. This technique has been used largely in the comparison of differential expression between developmental stages, responses to stress and between wild-type and mutant to ascertain potential gene targets and networks[130,291,292].

RNA-Seq takes advantage of the latest advances in sequencing technology. RNA sampled from the tissues of interest are reverse transcribed to cDNA with the

incorporation of sequencing adaptors at the ends of each nucleotide sequence. cDNA fragments are anchored to a solid platform and then re-sequenced. Following sequencing the reads are mapped back to a consensus sequence, with the number of reads at a locus proportional to the level of gene expression[293].

Although both microarray and RNA-seq have the advantage of genome-wide gene expression information there are disadvantages to the two techniques. In particular, micro-array will only provide gene expression information for those regions of the genome represented by the oligonucleotide probes present on the chip which in some cases can lead to gene expression going undetected, particularly if alternative splice variants are present within the sample. Whilst RNA-seq is not limited by the absence of complementary probes, it does require a large amount of post-sequence bio-informatics processing and an annotated reference DNA sequence against which sequence reads can be mapped. Quantitative RNA-seq experiments are further dependant on the read depth, the number of times a particular sequence is read during the sequencing process[294,295]. When studying individual loci, qRTPCR is recognised as the most accurate method to quantify gene expression in plant tissues and is often used as a means of validating the findings of RNA-seq and micro-array experiments.

In-situ hybridisation is another technique that is used to identify which specific tissue/cell-type within an organism a gene is expressed. Fluorescently labelled probes targeting genes of interest are applied to fixed tissue sections and observed using microscopy. Fluorescence is detected in the regions of the tissue where the gene of interest is expressed[296].

With respect to other known row-type genes, *VRS1*, *INT-C* and *VRS4*, previous studies in barley have looked at the levels of their expression using varying tissues and techniques.

INT-C, in barley, has not been localised to a specific tissue or cell-type but its maize orthologue *TB1* has been localised to axillary meristems, immature cob and early developing kernel with a role in lateral branching and plant architecture[281,297]. In rice, using a wild type promoter and β-glucoronidase fusion, *OsTB1* expression was observed in the axillary bud and the basal part of the shoot apical meristem supporting *OsTB1* as a negative regulator of lateral branching in rice[127]. In barley using microarrays, *INT-C* was found to show low expression with highest levels observed in the developing

inflorescence[291]. Similarly, qRTPCR analysis of *INT-C* in developing inflorescence and leaf tissue in the variety Bowman and the Bowman isoline BW421 (*int-c*.5), also observed increased levels of expression in the developing inflorescence compared to the leaf tissue[72]. There are two identified naturally occurring alleles of *INT-C* (*Int-c*.a (six-rowed) and *int-c*.b (two-rowed))[72] and at present it is not known whether these two alleles have different gene expression patterns and localisations.

Vrs1 is unique to barley, and expression is localised to tissue of the lateral spikelet primordia [298]. *Vrs1* expression was found to decrease with spike development, with the greatest levels of expression observed in the 1—4 mm developing inflorescence[78]. In a subsequent study, *Vrs1* expression was further localised to the lemma, palea, lodicule, pistil and rachillae of the lateral florets from inflorescence at the white anther stage of development[121]. Also, in this study, the expression of *Vrs1* in differing row-type induced mutants was investigated. Comparison was made between induced mutants and parental cultivars in non-isogenic backgrounds. The most significant reduction in the expression of *Vrs1* was observed in a *vrs4.l* mutant (19bp deletion mutant[130]) background concluding that *Vrs4* has a role in the transcription of *Vrs1*. In *hex-v*.44 (a premature stop codon mutant allele of v*rs1*[78]) the expression of *Vrs1* was significantly reduced compared to the wild-type two-rowed phenotype Bonus. Significant reduction of *Vrs1* expression was also detected in the *Vrs2.e* allele compared to parent cultivar, Svanhals. However, no significant differences in the expression of *Vrs1* between mutant and wild-type were detected in the Bonus derived mutant alleles *int-a*.1 (*vrs3*), and *int-c*.5[121].

VRS4 is an orthologue of the maize inflorescence architecture gene *RAMOSA2* (*RA2*). *VRS4* expression was first detected at the double ridge stage of inflorescence development, prior to the differentiation of lateral and central spikelets. Subsequently, *VRS4* was found to be most abundant in the lateral spikelet primordia but some expression was also detected within the central spikelet primordia. Expression levels of *Vrs4* increased from the triple mound to glume primordia stage but subsequently declined with inflorescence development[130]. Micro-array analysis of the differential expression of *Vrs4* mutant and wild-type backgrounds confirmed the reduced expression of *Vrs1* in a *vrs4* mutant background, but did not give any detail to difference in expression at the *Int-c* locus[130].

Although nothing is yet known about the expression of *Vrs3* in barley, in rice the orthologue JMJ706 was localised to the cell nucleus[202], consistent with its role as a histone demethylase. This study also suggested the LOB domain containing gene *Degenerate Hull 1* (DH1) and the MADS box transcription factor *OsMADS47* as regulatory targets of *JMJ706*[202]. A micro-array experiment, analysing the expression of 57,381 probe-sets across 39 developmental tissues of rice found the expression of JMJ706 (LOC_Os10g42690) to decrease with panicle development. The highest levels of expression were observed in the lemma and palea immediately before anthesis, and also leaf samples harvested when the panicle was <1 mm and 14 days post anthesis[299].

RNAseq expression values for the JMJ706 locus in The Rice Genome Annotation Project[147] suggest expression across all tissues sampled, with the highest levels of expression in the pre-emergent inflorescence, anther and pistil. The predicted maize orthologue of *VRS3,* GRMZM2G180086, was found to be constitutively expressed across all 60 tissues screened with a micro-array of 330,788 probes designed to 30,892 cDNA gene models as part of the Maize gene expression atlas[297].

5.1.1 ***Chapter Summary***

In this chapter, two methods were used to further explore the expression profiles of the known row-type genes *Vrs1*,*Vrs3, Vrs4* and *Int-c*. Firstly the Morex RNA-seq tissue expression atlas[57] was interrogated for their tissue specific expression profiles. Secondly, reciprocal qRTPCR analysis of the known row-type genes across their respective Bowman near isogenic mutants was used to identify regulatory interactions between the row-type loci.

5.2 *Materials and methods*

5.2.1 *Barley RNAseq 16 tissue gene expression atlas*

The barley RNA-seq atlas was developed from 8 developmental tissues isolated from the six-rowed spring barley cultivar Morex[57]. Subsequent to this a further 8 tissues were isolated and underwent RNA seq analysis[300]. Genomic DNA sequences of the respective Morex alleles of identified row-type genes: *vrs1*, *Int-c*, *Vrs3* and *Vrs4* were used to reciprocal BLAST against the Morex Genomic Sequence (version3) to establish genomic DNA sequence contigs. These contigs were subsequently reciprocally BLASTed against the BarleyGenes Barley RNAseq Database[300] to establish transcript expression profiles. Table 32 lists the alleles and respective sequence identifiers.

Table 32: The alleles and genomic sequence identifiers used to BLAST against the Morex WGS database and BarleyGenes RNAseq database.
The database identifier of the respective mRNA transcript expression profile is also detailed.

Allele	Source Genomic Sequence	Morex WGS (version 3) Genomic DNA contig	Expression Atlas mRNA Transcript identifier
vrs1.a[78]	Genbank accession: AB489127	contig_135757	JLOC1_3180
Vrs3.x (Chapter 3)	Resequencing (Chapter 3)	contig_5669	JLOC1_46139
Vrs4 (haplotype 1)[130]	Genbank accession: KC854546.1	contig_2547112	JLOC1_23621
Int-c.a[72]	Ramsay *et al*.[72]	contig_5747	unmapped*

*****The transcripts of *Int-c*.a were not mapped in the expression atlas database as their level was below the threshold. The raw RNAseq reads for this locus were revisited and the expression level determined.

5.2.2 *qRTPCR of identified row-type genes*

5.2.2.1 *Germplasm*

To minimise background genetic effects, all row-type mutants used in the qRTPCR study were from the Bowman near isogenic line collection[114] (Table 33).

Table 33: The Bowman near isogenic line germplasm used in the qRTPCR study.

Bowman NIL	Allele	Mutant polymorphism	Mutant Donor source
BW898	*vrs1*.a	1 bp deletion[78]	Glenn
BW901	*vrs2.e*	N/A*	Svanhals
BW902	*vrs3*.f	1 bp deletion	Hakata 2
BW903	*vrs4*.k	1bp deletion[130]	MFB-2
BW421	*int-c*.5	1 bp deletion[72]	Bonus
Bowman	-	-	-

*The gene underlying the *Vrs2* locus is yet to be identified

5.2.2.2 *Glasshouse design and growth conditions*

Individual grain were sown into multicellular trays (24 cells, 6 cm × 6 cm) containing a compost/Intercept granular insecticide mix suitable for growing cereals. Plants were grown in a glasshouse at 18^{o}C under long day conditions.

The trial was comprised of three biological replicates and three growth stages. Within each biological rep, each growth stage comprised a multicellular tray of 24 plants of one of the 6 genotypes listed in Table 33. The sowing of biological replicates was staggered, with 3 day intervals between sowings (rep 1 sown: 26/06/12). The experiment was designed as a randomised complete block design, with each rep occupying a 3 tray by 6 tray block on the glasshouse bench. A border of guard plants was sown around the edge of the trial to minimise any edge effect.

5.2.2.3 *Tissue harvest*

The leaf sheath of the main stem of individual plants was dissected using fine forceps to expose the developing inflorescence. The developing inflorescence was harvested at three different growth stages: 1 mm, 2—3 mm and 10 mm, which corresponded to 5 day intervals between growth stages at the ambient glasshouse temperature. Inflorescence tissues were harvested using fine forceps followed by immediate freezing in liquid nitrogen. Frozen inflorescence tissue was stored at -80^{o}C until RNA extraction.

5.2.2.4 *RNA isolation, quality check and cDNA synthesis*

Approximately 50 mg of frozen inflorescence tissues were ground to a fine powder in micro-centrifuge tubes by sterile micro-pestle. Micro-centrifuge tubes were suspended in a bath of liquid nitrogen to ensure tissue remained frozen during the grinding process.

RNA extraction proceeded using the QIAGEN RNeasy plant mini extraction kit (Catalogue Number: 74904) with RNA eluted into a 50 µl volume. This was followed by a further DNAse treatment using DNA-free™ DNase (Invitrogen) as per the manufacturers guidelines. RNA quality was assessed by Bioanalyzer with all RNA having a RNA integrity number (RIN) value greater than 8. 2 mg of RNA was converted to cDNA using Ready-To-Go You-Prime First-Strand Beads (GE Healthcare Lifesciences). RNA was incubated at 65^{o}C for 10 minutes, chilled on ice for 2 minutes, 1.13 µl 50 µM oligoDT, 0.37 µl 50 µM random hexamers and reverse transcription beads were then mixed with the RNA. The reverse transcription reaction proceeded with incubation at room-temperature for 1 minute, followed by incubation at 37^{o}C for 1hr. cDNA was diluted to a final volume of 100 µl.

5.2.2.5 *Gene Targets and Primer Design*

Barley orthologues for *Act2* (*Actin2*) and *PDF2* (*Protodermal Factor 2*) were used as reference genes, having been previously determined as stably expressed across a range of tissues of Arabidopsis[301] and Barley (Calixto *et al.* unpublished).

qRTPCR assays were designed using the Roche UPL design centre to *Vrs1*, *Vrs3*, *Vrs4* and *Int-c* (https: //lifescience.roche.com/shop/CategoryDisplay?identifier=Universal+Probe+Library) (Table 34) and, where possible, primers/assays were designed to be intron spanning. The positions of primers relative to gene structure and causal polymorphism can be found in Appendix 23.

To ascertain specificity, amplicons and primers were BLASTed against the Morex genomic barley sequence[57]. Additionally, primers were tested for specificity by the presence of a single band of the appropriate size following PCR and agarose gel electrophoresis.

Table 34: qRTPCR Primers and respective Universal Probe Library Probes

Gene	Forward Primer	Reverse Primer	UPL probe
VRS1	CCCATAAAATAGCCGAGATAGC	AGGTTTCTGCCGATCTTGAA	#70 (04688937001)
VRS3	CACTTTCTTTATGAGTGGACGAAA	CAGAAGAGATTTCACGCCAGA	#101 (04692195001)
VRS4	GTGAACGCCATTAGCACCAT	GTGATCCATCCCAATGCTCT	#77 (04689003001)
INT-C	ACCATTCCTCCCCTCCATT	GCACCGGCACCGGCACAGAGGTAG	#31 (04687647001)
Act2	GCGAGTTGTCTGGGTCTTCT	ACATGGCAAGGACTTGAGAAA	#129 (04693655001)
Pdf2	TGGGTGCAGAAATAACATGC	ATGTTCGGCACTCTGTCCTT	#7 (4685059001)

As the gene underlying *VRS2* remained unidentified it was not possible to include *VRS2* as a target in the qRTPCR analysis. However, the *vrs2* Bowman isoline was used for RNA extraction to observe any effects the *vrs2* mutation had on the other identified row-type loci.

5.2.2.6 *Primer efficiency determination*

To determine PCR amplification efficiency, primer pairs were tested over a 1:4 cDNA serial dilution series. Primers with an R^2 of >0.99 and an efficiency of between 95% and 105% were deemed suitable for qRTPCR (Appendix 24).

5.2.2.7 *qRTPCR reaction conditions*

All reactions were carried out on a StepOnePlus real-time PCR machine (Applied Biosystems).

Individual PCR reaction volumes totalled 25µl, comprising of 80ng cDNA, 2X Universal probe library Mastermix +ROX (04913957001), 20 µM forward and reverse primers, 10 µM Universal Probe Library hydrolysis probe.

PCR conditions were as follows:

95°C —10 minutes
40 cycles of:
95°C—15 s
60°C—1 minute

5.2.2.8 *PCR plate setup*

Each 96 well PCR plate comprised cDNA samples from 1 biological replicate of 1 inflorescence growth stage. Reactions on all plates comprised the amplification of one row-type gene (*VRS1, VRS3, VRS4* or *INT-C*) and the two-reference genes across three technical replicates.

5.2.2.9 *Calculation of relative expression levels*

The Pfaffl method[302] was used to calculate expression level of each gene.

$$RQ = E^{Ct(Bowman) - Ct\ (Bowman\ NIL)}$$

Where RQ is the relative quantification, E the efficiency of the primer pair and Ct, the average PCR cycle number across the three technical replicates at which the reaction reached the threshold.

RQ values were normalised to the geometric mean of the reference genes, *Pdf2* and *Act2*[303]:

$$NRQ = \frac{RQ\ (Gene\ of\ interest)}{\sqrt[2]{RQ(Pdf2)\ \times\ RQ\ (Act2)}}$$

where, NRQ is the normalised relative quantification and RQ the relative quantification.

General ANOVA analysis (Genstat version 14[141]) was used to determine significant differences between means. Due to the exponential scale of qRTPCR data, prior to ANOVA analysis, all NRQ were log transformed to comply with rules of normality. Fisher's protected LSD multiple comparison test was used to identify significant differences between means post ANOVA.

5.3 *Results*

5.3.1 ***RNAseq tissue specific transcription of row type genes***

Little is known about the spatial expression profile of the known barley row-type genes, *Vrs1*, *Vrs3*, *Vrs4* and *Int-c*. In the case of *Vrs1* and *Vrs4*, localisation was limited to tissues of the developing inflorescence[78,120,130]. The expression of barley row-type genes was analysed in the available barley RNA-seq database across a range of developmental stages to compare their expression and to identify any regulatory interactions between the genes.

The RNA-seq tissue expression atlas of barley comprises 16 tissues sampled at varying stages of growth and development of the six-rowed cultivar Morex (*vrs1*.a, *Int-c*.a).

Table 35: Description of the tissues that are included in the RNA-seq tissue expression atlas

Tissue	Description
Embryo	Dissected from grain, 4 days post germination
Seedling Root	Harvested from seedlings 17 days after planting Primary Roots 10—15 cm
Seedling shoot	Shoots harvested from seedlings 17 days after planting
Inflorescence (5 mm)	Inflorescence harvested 30 days after planting (5 mm)
Inflorescence (10—15 mm)	Inflorescence harvested 50 days after planting (10—15 mm)
3rd Internode	3rd Internode harvested 42 days after planting
Developing grain (5DPA)	Caryopsis harvested 5 days post anthesis
Developing grain (15DPA)	Caryopsis harvested 15 days post anthesis
Etiolated Seedling	Seedling grown in dark, harvested 10 days after planting
Lemma	Harvested 42 days after planting (yellow anther stage)
Lodicule (including stigma)	Harvested 42 days after planting (yellow anther stage), Lodicules at maximum swelling pre-anthesis
Palea	Harvested 42 days after planting (yellow anther stage)
Peeled Epidermis	Harvested 16 days after planting
Rachis	Harvested 35 days after planting, pre-anthesis
Mature Root	Harvested from 28 days after planting, Primary Roots >30 cm
Senescing Leaf	Harvested 32 days after planting

.

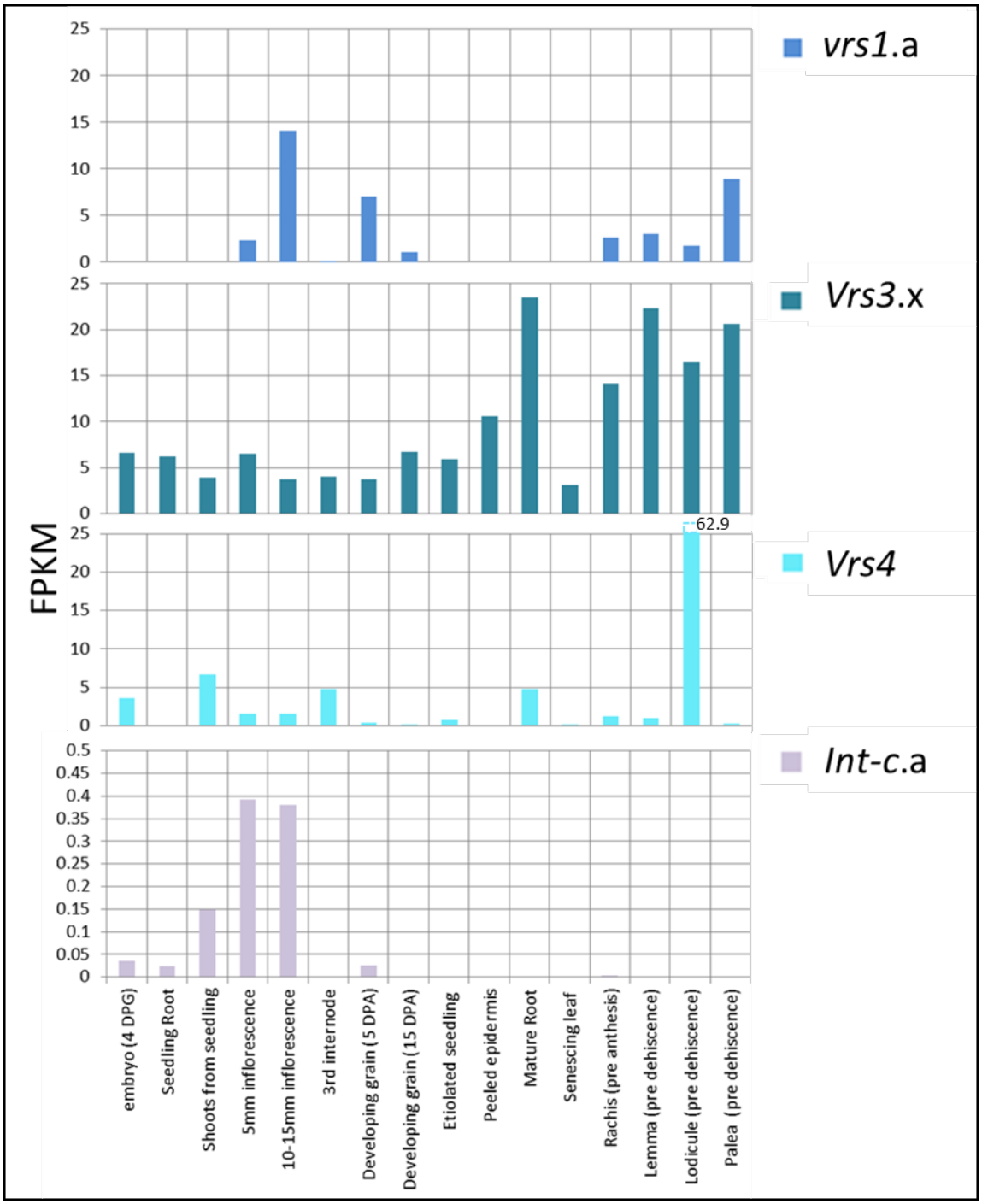


Figure 52: The differential expression profiles of the four identified row-type loci *vrs1, Vrs3, Vrs4 and Int-c* across 16 different tissues of six-rowed cultivar Morex as determined from RNAseq data.
The specific alleles of the loci are shown on the right of the graphs. Expression data is shown as FPKM (Fragments Per Kilobase of exon per Million fragments mapped). For reference of expression levels, expression of constitutively expressed genes *PDF2* and *ACT2* varied between 78—230 and 124 — 591FPKM respectively.

Of the four row-type genes, *Vrs3* displayed the most diverse expression profile, with expression detected across all 16 tissues but, in particular, mature root and the specific inflorescence tissues gave the highest levels of expression. Conversely, the expression patterns of *vrs1* and *Int-c* were restricted to specific tissues. *vrs1* expression was only detected in tissues related to inflorescence architecture and early stages of developing grain. *Vrs4* expression, whilst covering a diverse range of tissues from embryo to mature root was generally at a lower level compared to *Vrs3*. *Vrs4* showed the highest level of expression in the lodicule (+stigma) tissue, with an FPKM of 62.9, which was nearly 4x higher than the expression of *Vrs3* and 35x higher than the expression of *Vrs1*, indicating a very specific upregulation of expression of *Vrs4* in this tissue.

Int-c expression was by far the lowest over all the tissues and showed its greatest level of expression in the developing inflorescence. However, the levels of *Int-c* expression in the developing inflorescence were only 6% the level of *Vrs3* and 20% the level of *Vrs4* in the 5 mm developing inflorescence and 3% the level of *vrs1* in the 10—15 mm developing inflorescence.

Vrs3 and *Vrs4* were expressed (FPKM >4) at the earliest stage of plant development in the germinating embryo both continued to show similar expression levels in the seedling shoot. Some *Int-c* expression was detected in the shoot grown under normal light conditions but no *Int-c* expression was detected in etiolated seedling.

Vrs3 and *Vrs4* were also detected in the third internode tissue suggesting a more general role for these genes in plant architecture. However, only *Vrs3* was detected in the epidermal layer of the leaf and in the senescing leaf.

Vrs3, *vrs1* and *Int-c* showed differences in expression levels between the two developing inflorescence tissues. Expression of *Int-c* decreased slightly (approximately 3%) between the 5 mm and 10—15 mm inflorescence tissue while *Vrs3* showed a proportionally greater decrease of 42% and *vrs1* which showed a six-fold increase in expression in the later stage. This is in contrast to *Vrs4* which showed low levels of expression in both inflorescence tissues and no change in the level of expression. At the latter stages of floral development, pre-anthesis, *vrs1*, *Vrs4* and *Vrs3* were all expressed at varying levels.

Int-c, *Vrs3* and *Vrs4* are all expressed in varying stages of root development, *Int-c* expression was restricted to seedling root development, *Vrs4* to mature root but *Vrs3*

showed a more global root expression in both seedling and mature roots. This expression has not been described in barley previously and suggests possible roles in root architecture in addition to the previously documented role in inflorescence architecture. Similarly, the diverse expression pattern of *Vrs3* suggests a role not limited to lateral spikelet development; however as non-inflorescence related phenotypes are yet to be identified in *vrs3* it could suggest that perhaps *Vrs3* has a constitutive function and has greater importance in higher expressed tissues.

5.3.2 ***qRTPCR of row-type genes across developing inflorescence tissue***

Mutants of barley row-type genes show an altered phenotype in the barley inflorescence. In this study a qRTPCR approach was used to compare the expression of the row-type genes *VRS1*, *VRS3*, *VRS4* and *INT-C* across the Bowman near isogenic lines BW898 (*vrs1*.a), BW901 (*vrs2*.e), BW902 (*vrs3*.f), BW903 (*vrs4*.k) and BW421 (*int-c*.5) relative to their expression in wild-type Bowman.

5.3.2.1 *Growth stages of harvested tissues*

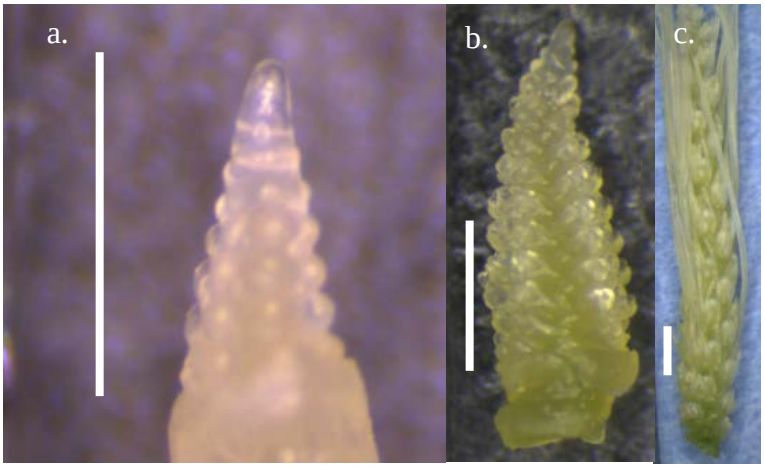


Figure 53: Representative developing infloresecence from Bowman illustrating the three growth stages at which inflorescence tissue was harvested prior to qRTPCR
a) approximately 1 mm b) 2—3 mm and c) 10 mm. In all cases, white bars indicate 1 mm.

Developing inflorescence were harvested at three growth stages to reflect the different stages of development (Figure 53). The approximate growth stage of the inflorescence was determined by the size of the inflorescence and stage of the spikelets in the central region of the inflorescence; the precise developmental stage of the inflorescence was difficult to determine given the graduated nature of spikelet development across the inflorescence but in general growth stage 1 was inflorescence at the triple mound stage, growth stage 2 the lemma primordia and growth stage 3 late awn primordium.

5.3.2.2 *Relative quantification of row-type gene expression*

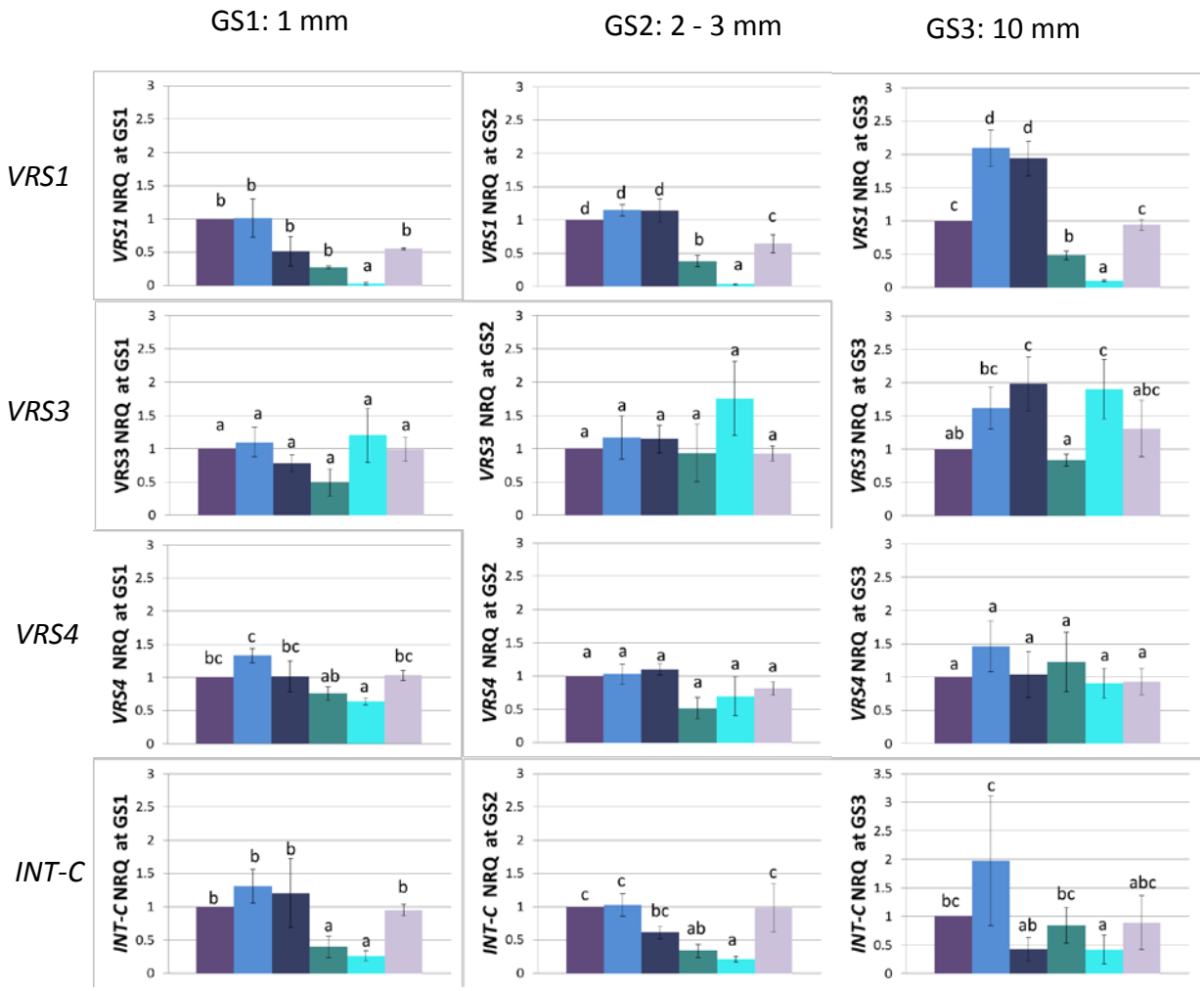


Figure 54: NRQ of row-type genes at three growth stages of the developing inflorescence, in reciprocal mutant near-isogenic backgrounds

Comparison of the normalised relative quantification (NRQ) of genes known to affect row-type, at three different stages of the developing inflorescence: GS1: 1 mm, GS2: 2—3 mm and GS3: 10 mm.

	Bowman
	BW898(*vrs1*.a)
	BW901(*vrs2*.e)
	BW902(*vrs3*.f)
	BW903(*vrs4*.k)
	BW421(*int-c*.5)

a) NRQ of *VRS1* b)NRQ of *VRS3*, c)NRQ of *VRS4* d)NRQ of *INT-C*.

Values shown are the arithmetic mean. All values are relative to the expression of Bowman and normalised to the expression of reference genes Act2 and Pdf2. Error bars are the standard error. Significant differences were determined by ANOVA analysis of the Log transformed means (Appendix 25).

5.3.2.3 *Vrs1* (Figure 54a)

Across all three growth stages, *Vrs1* expression was reduced to virtually zero expression in a v*rs4* mutant background compared to Bowman supporting an interaction between *Vrs4* and *Vrs1* expression. This effect is not reciprocated in the *Vrs4* expression in a *vrs1* mutant background (Figure 54c). At growth stage 2 and 3 significant 2—3 fold reductions in *Vrs1* were observed in the *vrs3* mutant background suggesting that *Vrs3* also regulates the level of *Vrs1* expression. *Vrs1* expression was more stable in an *int-c* mutant background although a significant reduction in expression of *Vrs1* was observed at growth stage 2.

At both growth stages the reduction in expression of *Vrs1* in the *vrs4* mutant background was significantly greater than the reduction in expression in a *vrs3* mutant background.

Both vrs1 and *vrs2* mutant backgrounds showed a significant increased expression of *VRS1* compared to wild-type Bowman at growth stage 3 indicating an induction of expression in response to lower levels of these proteins at the later growth stage. At growth stage 3 lateral spikelets in *vrs1* and *vrs2* mutants show enhanced development, compared to the wild-type lateral spikelets, with lemma awns formed. Whilst the function of *VRS2* remains yet to be determined, *Vrs1* is known to supress the development of lateral spikelets, therefore detection of its absence in the mutant could be resulting in promotion of enhanced expression and therefore the production of increased levels of mRNA.

5.3.2.4 *Vrs3 (*Figure 54b*)*

At growth stages 1 and 2 no significant difference in *Vrs3* expression was detected across all the mutant genetic backgrounds. However, at growth stage 3, both *vrs4* and *vrs2* mutant backgrounds showed significant two-fold increase in the levels of *Vrs3* expression compared to Bowman.

5.3.2.5 *Vrs4* (Figure 54c)

In contrast to the effect the *vrs4* mutant background had on the expression of *VRS1*, *VRS3* and *Int-c, Vrs4* expression was largely unaffected by the different mutant backgrounds. *Vrs4* expression was only significantly reduced compared to Bowman in a *vrs4* mutant background at growth stage 1. At growth stages 2 and 3 there was no significant difference in expression in the *vrs4* mutant background.

5.3.2.6 *Int-c* (Figure 54d)

Like the RNA-seq data, the expression levels of *Int-c* in the Bowman NILs were very low. At growth stage 3 in particular the standard errors in expression were large reflecting the very low levels of expression of *Int-c* and making consistent reliable detection difficult. Nevertheless, at growth stages 1 and 2 significant reduction in the expression of *Int-c* was observed in both *vrs3* and *vrs4* mutant backgrounds which mimics *Vrs1* changes in expression (Figure 54a). However, at growth stage 3 no significant difference in *Int-c* expression was detected across all genotypes, unlike the induction of *Vrs1* in a *vrs1* and *vrs2* mutant background (Figure 54a).

5.4 *Discussion*

This study of barley row-type gene expression has highlighted the complex regulation of these genes and their expression across different barley tissues, in particular, barley inflorescence development.

The barley RNA-seq tissue expression atlas revealed a constitutive expression pattern of *Vrs3* and to a lesser extent *Vrs4* across a range of developmental tissues and a considerably more tissue specific expression pattern for the six-rowed alleles of *vrs1* and *Int-c*. This suggests that *Vrs3* and *Vrs4* function is not limited to the inflorescence development pathway but may have more general roles across many different tissues. *Vrs3* expression was induced 3—5 fold in mature root rachis, lemma, lodicule and palea compared to the other tissues including developing inflorescence and *Vrs4* was highly expressed in the lodicule. It remains to be determined whether *vrs3* and *vrs4* mutant phenotypes show other tissue specific phenotypes that correspond to these expression patterns. For example, *Vrs3* was highly expressed in mature roots but *vrs3* mutants have not previously been screened phenotypically for differences in root architecture in barley.

Although not previously documented in barley roots, orthologues of *Int-c* in both maize (*ZmTb1*) and rice (*FC1*) have been associated with root architecture. *FC1* expression has been detected at a high level in the crown root tip, through in-situ hybridisation, but no obvious root phenotype was detected in the *fc1* mutant[304]. In maize, a dosage response has been reported with respect to the mutant *tb1-ref* allele; homozygous *tb1-ref* was found to produce significantly more lateral root branches compared to the heterozygous *tb1-ref* and wild-type *Tb1* alleles, there was also an associated increase in above ground branching in the homozygous *tb1-ref* mutant[305]. This would suggest that *Int-c* may have a role in controlling root out-growth in addition to the previously documented control of shoots. With recent improvements in methods for root screening including scanning[306] and transparent soil[307] it should be possible to identify any root architectural phenotypic differences which may provide evidence for a role of *Vrs3* and *INT-C* in barley root tissues. In this case both natural six-rowed and two-rowed alleles of *INT-C* should be considered as it may be that the expression of natural two-rowed alleles may differ in root expression pattern compared to the six-rowed allele of *Int-c* in the Morex RNAseq expression analysis.

RNA-seq data did not show high levels of *VRS3* gene expression in the developing inflorescence compared to other tissues, however qRTPCR expression analysis of three stages of developing inflorescence revealed that *VRS3* expression was induced when other row-type genes were mutated. This supports that *VRS3* expression responds to changes in the row-type pathway and plays a role in the inflorescence development pathway.

VRS1 expression was specific to developing inflorescence tissue but also showed expression levels within developing grain and inflorescence organs. *INT-C* was expressed at very low levels in all tissues but showed its highest levels within developing inflorescence consistent with its role in this tissue[72]. This tissue specificity was found in six-rowed Morex and a comparable RNA-seq study of tissue expression in a two-rowed barley cultivar would help confirm whether these tissue specific patterns of expression observed in the case of *INT-C* and *VRS1* are a general feature of the locus or due to the six-rowed allele.

The overlapping expression profiles and consistent down-regulation of both *Int-c* and *Vrs1* in *vrs3* and *vrs4* mutant backgrounds suggests that *Vrs3* and *Vrs4* both have a role in controlling the expression of *Int-c* and *Vrs1* in the developing inflorescence. The down-regulation of *Vrs1* in a *vrs4* background was significantly greater than the down-regulation of *Vrs1* in a *vrs3* background, which may be interpreted as *Vrs4* having a more direct role in regulating the transcription of *Vrs1*, whereas the role of *Vrs3* is more of transcriptional enhancement, consistent with suggestions of the role of histone demethylases in other species[241].

In maize, the orthologue of *Vrs4*, *Ramosa2* (*Ra2*) has been previously implicated in the regulation of TCP transcription factors in inflorescence architecture. For example the expression of the class 2 TCP transcription factor *Bad1* (*Branched angle defect 1*, the mutant phenotype of which results in upright maize tassel branches), was found to be halved in a *ra2* mutant background[308] consistent with the finding of reduction of *Int-c* expression in a *vrs4* background in barley.

The downregulation of *Vrs1* in a *vrs4* mutant background is consistent with the earlier study of Sakuma *et al*. which identified a significant difference in expression of *Vrs1* in a *vrs4* background at the later white anther stage of inflorescence development[121]. However, in contrast, their study found no significant difference in the expression of

Vrs1 in the induced mutant *int-a*.1 (*vrs3* syn.) compared to the wild-type parent Bonus which contradicts the finding of a decreased level of expression in the BW902 (*vrs3*.f) genotype at growth stages 2 and 3. Through the identification of *VRS3*, it is now known that the *int-a*.1 mutation is the result of a 2 bp deletion at the 3' end of *Vrs3*; in comparison the mutation used in this study, *vrs3*.f, is a deletion at the 5' end of *Vrs3* (Table 14) and is a phenotypically stronger mutant (section 3.3.5.1) which may help explain the differential results in the expression of *Vrs1*. It is also possible, due to the non-targeted nature of induced mutagenesis, other background mutagenic effects may influence the expression levels in the *int-a*.1 mutant. In the experiments described in this chapter these variable expression effects were minimised by utilising near isogenic lines.

At growth stage 2, in the developing inflorescence the expression of *Vrs1* was decreased in an *int-c*.5 mutant background, which was not observed in the earlier or later growth stages. The RNA-seq tissue expression profile suggested that *vrs1* expression increased with inflorescence development and *Int-c* expression decreased slightly. This in combination with the qRTPCR result suggests that *Int-c* may have a role in the regulation of *VRS1* but that its role may occur in the early stage of inflorescence development. Inflorescence tissue used by Sakuma *et al*. was at a later growth stage (white anther), a stage at which the expression of *Int-c* is not expected and therefore unlikely to control the expression of *Vrs1*. This may help explain the discrepancy between the findings in the qRTPCR study presented here and the earlier study by Sakuma *et al*. which suggested no significant difference in expression of *Vrs1* in an *int-c*.5 background.

Vrs4 showed exceptionally high levels of expression in the lodicule including stigma tissue immediately prior to anthesis (yellow anther stage), this expression is consistent with findings in the maize tissue expression analysis which identified the highest levels of expression of the *Vrs4* orthologue, *Ramosa 2*, to be in the meiotic tassel, pre-anthesis[297]. The function of the lodicule is to swell prior to anthesis, creating space in the spikelet to enable pollen to reach the stigma. In the barley RNAseq tissue atlas, the lodicule tissue was harvested pre-anthesis, at the yellow anther stage, when the lodicule would be nearly maximally swollen; it could therefore suggest that *Vrs4* has a role in regulating the rapid cell proliferation required in the swelling of the lodicule. A similar

role of *Ramosa 2* in regulating cell proliferation in the maize tassel pulvinus which swells to control maize tassel branch angle at maturity has previously been reported[308].

The qRTPCR results presented here suggest that *Vrs3* and *Vrs4* have a role in the regulation of *Vrs1* and *Int-c* but what remains unclear are the mechanisms of regulation of *Vrs3* and *Vrs4*. It is also not clear that the significant increases in *Vrs3* expression at growth stage 3 in the *vrs4* and *vrs2* backgrounds are due directly to regulation by these genes particularly as one of these genes, underlying *VRS2,* is yet to be identified. Previous reports have suggested a role of micro-RNA 396d directed regulation of the rice *VRS3* orthologue, *JMJ706*, mediated through *GRF* (Growth response factor) genes. Interaction between the Os*GRF* and *JMJ706* was found to be via a cis-acting GARE (gibberellic acid response element) 5' of the transcription start site[309]. An initial assessment of the barley *Vrs3* sequence using the New PLACE database[310] did not identify a GARE element in the 5'UTR but did identify an element in the 3'UTR. Alternatively, differential expression of the *Int-c* orthologue of maize, *Tb1*, has been shown to be mediated through an upstream enhancer locus[126]. It is yet to be identified whether a similar promoter and enhancer element are present in barley but it is possible that *VRS4* and *VRS3* may target such promoter elements rather than interacting with *Int-c* itself, but again further experiments would be required to determine any interaction. Promoter analysis in the mutant backgrounds, yeast two-hybrid protein-protein interaction analysis and expression analysis from double and triple mutant combinations of the Bowman isolines may help resolve the interaction and regulation between these genes.

At growth stage 3 both v*rs1* and *vrs2* showed significantly greater expression of *VRS1* compared to Bowman; the reason for this may be a regulatory response to the reduction of VRS1 and VRS2 in their corresponding mutants. However, it may also be due to the v*rs1* and *vrs2* genotypes producing more lateral spikelet tissue compared to Bowman and therefore the RNA sample will contain proportionally more cells from the lateral spikelets compared to two-rowed Bowman leading to the observed increase in expression of *VRS1*.

Gene expression analysis by RNA-seq and qRTPCR has revealed both tissue specific and constitutive expression in 4 barley row-type genes. Analysis of expression of these genes in their mutant background shows a complex interaction and response to the altered gene levels. *VRS3* expression showed background expression across many

tissues and was induced in roots and pre-dehiscence infloresence tissues. In the mutant *vrs3* background, predicted loss of *VRS3* function led to the reduction of *VRS1* and *INT-C* expression in developing inflorescence tissue, supporting the lateral spikelet phenotype. Micro-array or further RNA-seq analysis of these mutants will reveal global changes in gene expression in these phenotypically altered tissues and help identify genes altered in their expression by these row-type genes.

6. Conclusions and Future Perspectives

This study set out to identify and characterise the gene underlying the spike morphology mutant of barley *vrs3*. It also aimed to investigate the interaction of *VRS3* with other known row-type loci and assess the potential of *vrs3* as a means of improving grain uniformity within six-rowed barley cultivars.

VRS3 was identified as a JmjC histone demethylase, tightly linked to SNP 11_21217, positioned at 49.17 cM (IBGSC consensus map) on chromosome 1H. The nature of JmjC histone demethylases suggests that *Vrs3* may have multiple target loci within the barley genome. Whilst direct targets are yet to be identified, expression analysis suggested a possible role of *Vrs3* in the regulation of both *VRS1* and *INT-C*. An indirect role of *Vrs3* with respect to row-type is also supported by the finding of both *Vrs3*.x and *Vrs3*.w wild-type alleles in both two-rowed and six-rowed cultivars, not complete linkage between row-type and allele as is the case with the *Vrs1* and *Int-c* loci. However, the regulation of *Vrs1* by *Vrs3* appears to be not as determinate as that of the *Vrs4* gene *HvRa2*, where mutation of *vrs4* results in highly suppressed expression of *Vrs1*[130]. This would suggest *Vrs3* functions to alter transcription efficiency rather than as an "on/off switch".

Identifying the regulatory targets of *Vrs3* is key to furthering our understanding of the pathways in which *Vrs3* functions. Some targets could potentially be identified through the pleiotropic phenotypes observed; the awned palea[246] mutant represents one such target. Another possible regulatory candidate is *HvHox2*; whilst phylogenetic analysis suggested early evolutionary origins for *Vrs3*, this is in contrast to the duplication of *HvHox2* to form *Vrs1*[121]. In this study *Vrs3* was implicated in the regulation of *Vrs1*, it would therefore seem logical that *HvHox2* could also be a regulatory target of *Vrs3*. In addition to the focussing on individual key candidate loci, a more genome-wide approach to target identification could be taken. An RNAseq approach could be used to identified genes differentially expressed between wild-type and mutant in the differentiating inflorescence. An alternative approach could be to use Chipseq, with immunoprecipitation of the *Vrs3* target histone tail modification prior to resequencing. However, prior to this the histone tail modifications which *Vrs3* targets needs to be confirmed. Whilst the rice orthologue JMJ706 was found to target H3K9me2 and H3K9me3[202] it cannot be assumed that *Vrs3* targets the same residues, so an assay of *Vrs3* demethylase activity would be required to test this.

The *vrs3* allelic series[136], as well as providing an excellent means of confirming the gene underlying *VRS3*, now represent a major resource for the in-depth study of JmjC histone demethylases. The diverse collection of mutations identified within all three functional domains should provide insight into the mechanisms of target recognition by JmjC histone demethylases. However, this study has suggested that genetic background has a strong influence on the *vrs3* phenotypes observed; therefore creating a series of near-isogenic line collections of the *vrs3* alleles would enhance studies by minimising the effects of genetic background

Whilst there is no evidence of *vrs3* mutation significantly affecting time to spike emergence under long day conditions, there is evidence that time to reach spike emergence affects the *vrs3* spike phenotype observed. It would therefore be interesting to investigate allelic diversity at the known photoperiod[153,251,252] and earliness per-se[66,159] genes to see if any link can be established between the difference in *vrs3* phenotype observed and allelic diversity at these loci.

Phenotyping in this and previous studies of *vrs3*[132-134,267-269,311] has focussed primarily on the spike, however, consistent with finding in rice[299] and maize[297,305], there is evidence for *Vrs3* and *Int-c* expression within root tissue. This could suggest that root and spike development share elements of regulatory control or alternatively could a change in root phenotype drive the change in spike morphology observed?

A characteristic feature of the *vrs3* phenotype is the six-rowed phenotype in the upper portion of the spike and two-rowed phenotype at the base of the spike[136]. Within wild-type barley, it is established that barley spikelet development proceeds acropetally[23]. The distinct phenotypic development between the lateral spikelets in the lower portion of the spike in *vrs3* mutants compared to the upper would suggest differences in developmental regulation. This is also supported by the finding of increased grain development in the lower portion of the spike when *vrs3* is in combination with the six-rowed allele of *INT-C*, *Int-c*.a. but not to the same extent as observed in the full six-rowed phenotype when in combination with heterozygous or homozygous six-rowed *vrs1*. To determine the extent of differences in development regulation between the upper, middle and lower lateral spikelets, micro-dissection could be used to isolate specific developing spikelets followed by RNA-seq to establish differential gene expression.

Across the *VRS3* mapping populations and the *vrs3* allelic series, the *vrs3* phenotype showed high levels of plasticity suggesting a role for *Vrs3* as an environmental sensor. The stimuli to which *VRS3* reacts remain yet to be determined but could provide valuable insight into potential mechanisms for securing yield stability across differing environments.

Whilst the barley RNAseq tissue expression atlas has shown that *Vrs3* is constitutively expressed across all the tissues analysed, it remains to be established whether *Vrs3* is expressed globally or confined to specific cell-types. Establishing this through in-situ hybridisation with inclusion of probes directed to the other row-type genes, *Vrs1*, *Int-c* and *Vrs4* would further establish any overlapping cell expression profiles.

Whilst the component parts of barley inflorescence development pathway are gradually being identified[72,78,130,181], how the various pieces fit together remains elusive. This study suggests a role of *Vrs3* as a probable mediator of expression of *Int-c* and *Vrs1*. However, the global expression pattern of *Vrs3* also suggests a role out with the inflorescence development pathway.

Lundqvist *et al*. had previously suggested that varying combinations of intermedium loci could result in an enhanced six-rowed spike phenotype[134]. The findings of this study support this; however, although enhancement in both lateral grain width and area was observed in treble homozygous six-rowed allele combinations of *vrs1, Int-c* and *vrs3*, this was only significantly better than the bi-allele combinations of *vrs1, Int-c* and *vrs1,vrs3* under glasshouse conditions. The addition of six-rowed alleles was associated with a reduction in tillering, so whilst the treble homozygous six-rowed allele combination could offer potential improvement in uniformity in six-rowed cultivars, it would likely be associated with reductions in yield per plant due to the reduction in tillering; although this may not be significant at commercial planting densities. A plot trial grown at commercial sowing density would be useful to further test this hypothesis.

The identification of *Vrs3* as a JmjC histone demethylase, the first to be identified through map based cloning in barley, offers not only the potential to understand further the regulation of the barley spike row-type phenotype and means of its improvement but also the regulation of histone demethylation, an area which is yet to be studied in detail within barley.

7. Reference List

1 Food and Agriculture Organization of the United Nations (2014) FAOSTAT

2 Newton,A.C. *et al.* (2011) Crops that feed the world 4. Barley: a resilient crop? Strengths and weaknesses in the context of food security. *Food Security* 3, 141-178

3 Blake,T. *et al.* (2011) Barley Feed uses and Quality Improvement. In *Barley: Production, Improvement and Uses* (Ullrich,S.E., ed), pp. 522-531, Wiley-Blackwell

4 Garstang,J.R. and Spink,J.H. (2011) Cultural Practices: Focus on Major Barley-Producing Regions. In *Barley: Production, Improvement and Uses* (Ullrich,S.E., ed), pp. 221-281, Wiley-Blackwell

5 Euromalt (2013) Euromalt

6 Maltsters Association of Great Britain (2015) UKMalt.com

7 Molina-Cano,J.L. *et al.* (1995) EFFECT OF GRAIN COMPOSITION ON WATER UPTAKE BY MALTING BARLEY: A GENETIC AND ENVIRONMENTAL STUDY. *Journal of the Institute of Brewing* 101, 79-83

8 Ellis,R.P. and Marshall,B. (1998) Growth, yield and grain quality of barley (*Hordeum vulgare* L.) in response to nitrogen uptake: II. Plant development and rate of germination. *Journal of Experimental Botany* 49, 1021-1029

9 Crescenzi,A.M. (1987) FACTORS GOVERNING THE MILLING OF MALT. *Journal of the Institute of Brewing* 93, 193-201

10 Samarah,N.H. (2005) Effects of drought stress on growth and yield of barley. *Agronomy for Sustainable Development* 25, 145-149

11 Saini,H.S. and Westgate,M.E. (1999) Reproductive Development in Grain Crops during Drought. *Advances in Agronomy* 68, 59-96

12 Kuczyñska,A. *et al.* (2013) Effects of the semi-dwarfing *sdw1*/*denso* gene in barley. *Journal of Applied Genetics* 54, 381-390

13 Houston,K. *et al.* (2013) Variation in the interaction between alleles of *HvAPETALA2* and microRNA172 determines the density of grains on the barley inflorescence. *Proceedings of the National Academy of Sciences of the United States of America* 110, 16675-16680

14 Jürgens,G. (1995) Axis formation in plant embryogenesis: Cues and clues. *Cell* 81, 467-470

15 Greb,T. *et al.* (2003) Molecular analysis of the *LATERAL SUPPRESSOR* gene in *Arabidopsis* reveals a conserved control mechanism for axillary meristem formation. *Genes & Development* 17, 1175-1187

16 Doebley,J. *et al.* (1995) *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* 141, 333-346

17 Vollbrecht,E. *et al.* (2005) Architecture of floral branch systems in maize and related grasses. *Nature* 436, 1119-1126

18 Liang,W. *et al.* (2014) Tillering and panicle branching genes in rice. *Gene* 537, 1-5

19 Tanaka,W. *et al.* (2013) Grass Meristems II: Inflorescence Architecture, Flower Development and Meristem Fate. *Plant and Cell Physiology* 54, 313-324

20 Wang,Y. and Li,J. (2008) Molecular Basis of Plant Architecture. *Annu. Rev. Plant Biol.* 59, 253-279

21 Nair,S.K. *et al.* (2010) Cleistogamous flowering in barley arises from the suppression of microRNA-guided *HvAP2* mRNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America* 107, 490-495

22 Zadoks,J.C. *et al.* (1974) A Decimal Code for the Growth Stages of Cereals. *Weed Research* 14, 415-421

23 Kirby,E.J.M. and Appleyard,M. (1981) *Cereal Development Guide*, National Agricultural Centre

24 Takahashi,R. *et al.* (1975) Basic Studies on Breeding Barley by the Use of Two-Rowed and Six-rowed Varietal Crosses. In *3rd International Barley Genetics Symposium* pp. 662-677

25 Cannell,R.Q. (1969) The Tillering Pattern in Barley Varieties .I. Production Survival and Contribution to Yield by Component Tillers. *Journal of Agricultural Science* 72, 405-422

26 Kirby,E.J.M. and Faris,D.G. (1972) The effect of plant density on tiller growth and morphology in barley. *The Journal of Agricultural Science* 78, 281-288

27 Cannell,R.Q. (1969) The Tillering Pattern in Barley Varieties 2. Effect of Temperature Light Intensity and Daylength on Frequency of Occurrence of Coleoptile Node and Second Tillers in Barley. *Journal of Agricultural Science* 72, 423-435

28 Gustafsson,Å. *et al.* (1969) A Proposed System of Symbols for Collection of Barley Mutants at Svalöv. *Hereditas-Genetiskt Arkiv* 62, 409-414

29 von Zitzewitz,J. *et al.* (2005) Molecular and Structural Characterization of Barley Vernalization Genes. *Plant Molecular Biology* 59, 449-467

30 Szücs,P. *et al.* (2009) An Integrated Resource for Barley Linkage Map and Malting Quality QTL Alignment. *The Plant Genome* 2, 134-140

31 Beaven,E.S. (1947) Barley—-50 years of Observation and Experiment, Duckworth

32 Martin,P. *et al.* (2009) New Markets and Supply Chains for Scottish Bere Barley. In *European Landraces: on-farm conservation, management and use* (Veteläinen,M.*et al.*, eds), Bioversity International

33 Institute of Brewing (1930) Monthly Review — Visit to Dr Beaven's Barley Nursery and Field Plots at Warminster. *Journal of the Institute of Brewing* 36, 399-420

34 Hind,H.L. (1930) OBSERVATIONS ON SOME MALTING AND BREWING TRIALS WITH A 6-ROWED WINTER BARLEY. *Journal of the Institute of Brewing* 36, 435-439

35 AHDB (2015) UK Winter Barley Recommended List 2015/16

36 Comadran,J. *et al.* (2011) Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in *Hordeum vulgare* in Mediterranean environments. *Theoretical and Applied Genetics* 122, 1363-1373

37 Lawlor,D.W. *et al.* (1981) Growth of spring barley under drought: crop development, photosynthesis, dry-matter accumulation and nutrient content. *Journal of Agricultural Science* 96, 167-186

38 van Oosterom,E.J. and Acevedo,E. (1992) Adaptation of barley (*Hordeum vulgare* L.) to harsh Mediterranean environments III. Plant ideotype and grain yield. *Euphytica* 62, 29-38

39 Manninen,O. and Nissilä,E. (1997) Genetic Diversity Among Finnish Six-Rowed Barley Cultivars Based on Pedigree Information and DNA Markers. *Hereditas* 126, 87-93

40 Forster,B.P. *et al.* (2004) Genotype and phenotype associations with drought tolerance in barley tested in North Africa. *Annals of Applied Biology* 144, 157-168

41 Harlan J.R (1968) On The Origin of Barley. In *Barley: Origin, Botany, Culture, Winterhardiness, Genetics, Utilization, Pests*, U.S. Department of Agriculture

42 Olesen,J.E. *et al.* (2011) Impacts and adaptation of European crop production systems to climate change. *European Journal of Agronomy* 34, 96-112

43 HAKALA,K. *et al.* (2012) Sensitivity of barley varieties to weather in Finland. *The Journal of Agricultural Science* 150, 145-160

44 Pourkheirandish,M. and Komatsuda,T. (2007) The importance of barley genetics and domestication in a global perspective. *Annals of Botany* 100, 999-1008

45 Komatsuda,T. and Mano,Y. (2002) Molecular mapping of the intermedium spike-c (*int-c*) and non-brittle rachis 1 (*btr1*) loci in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 105, 85-90

46 Pourkheirandish,M. *et al.* (2015) Evolution of the Grain Dispersal System in Barley. *Cell* 162, 527-539

47 De Candolle,A. (1908) *Origin of Cultivated Plants*, D. APPLETON AND COMPANY

48 Åberg,E. (1938) Hordeum agriocrithon, a wild six-rowed barley. Annals of the Agricultural College of Sweden 6, 159-212

49 Zohary,D. (1964) Spontaneous brittle six-row barleys, their nature and origin. In *The First International Barley Genetics Symposium* (1 edn) pp. 27-31

50 Tanno,K. and Takeda,K. (2004) On the origin of six-rowed barley with brittle rachis, *agriocrithon* [*Hordeum vulgare* ssp. *vulgare* f. *agriocrithon* (Åberg) Bowd.], based on a DNA marker closely linked to the *vrs1* (six-row gene) locus. *Theoretical and Applied Genetics* 110, 145-150

51 Murphy,P.J. *et al.* (1982) The Origin of 6-Rowed Wild Barley from the Western Himalaya. *Euphytica* 31, 183-192

52 Biffen,R.H. (1905) The Inheritance of Sterility in the Barleys. *Journal of Agricultural Science* 1, 250-257

53 Wiebe,G.A. (1968) Introduction of Barley into the New World. In *Barley: Origin, Botany, Culture, Winterhardiness, Genetics, Utilization, Pests* (US Dept of Agriculture, ed),

54 Molina-Cano,J.-L. *et al.* (2005) Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. *Theoretical and Applied Genetics* 110, 613-619

55 Dai,F. *et al.* (2012) Tibet is one of the centers of domestication of cultivated barley. *Proceedings of the National Academy of Sciences of the United States of America* 109, 16969-16973

56 Fischbeck,G. (1992) Barley Cultivar Development in Europe—Success in the past and possible changes in the future. In *Fourth International Barley Genetics Symposium* (2 edn) (Munck,L., ed), pp. 885-901

57 The International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491, 711-716

58 International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436, 793-800

59 The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796-815

60 Kleinhofs,A. *et al.* (1993) A Molecular, Isozyme and Morphological Map of the Barley (*Hordeum vulgare*) Genome. *Theoretical and Applied Genetics* 86, 705-712

61 Graner,A. *et al.* (1991) Construction of An RFLP Map of Barley. *Theoretical and Applied Genetics* 83, 250-256

62 Ramsay,L. *et al.* (2000) A Simple Sequence Repeat-Based Linkage Map of Barley. *Genetics* 156, 1997-2005

63 Mascher,M. *et al.* (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *The Plant Journal* 76, 718-727

64 LGC Genomics (2015) KASP Genotyping Chemistry

65 Close,T.J. *et al.* (2009) Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10, 582

66 Comadran,J. *et al.* (2012) Natural variation in a homolog of Antirrhinum *CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. *Nature Genetics* 44, 1388-1392

67 Muñoz-Amatriaín,M. *et al.* (2014) The USDA Barley Core Collection: Genetic Diversity, Population Structure, and Potential for Genome-Wide Association Studies. *Plos One* 9, e94688

68 Moore,G. *et al.* (1995) Cereal Genome Evolution — Grasses, Line Up and Form A Circle. *Current Biology* 5, 737-739

69 Krattinger,S. *et al.* (2009) Map-Based Cloning of Genes in Triticeae (Wheat and Barley). In *Genetics and Genomics of the Triticeae* (Muehlbauer,G.J. and Feuillet,C., eds), pp. 337-357, Springer US

70 Engler,F.W. *et al.* (2003) Locating Sequence on FPC Maps and Selecting a Minimal Tiling Path. *Genome Research* 13, 2152-2163

71 Whipple,C.J. *et al.* (2010) A Conserved Mechanism of Bract Suppression in the Grass Family. *Plant Cell* 22, 565-578

72 Ramsay,L. *et al.* (2011) *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED 1*. *Nature Genetics* 43, 169-172

73 Mayer,K.F.X. *et al.* (2011) Unlocking the Barley Genome by Chromosomal and Comparative Genomics. *Plant Cell* 23, 1249-1263

74 Schulte,D. *et al.* (2011) BAC library resources for map-based cloning and physical map construction in barley (*Hordeum vulgare* L.). *BMC Genomics* 12, 247

75 Matsumoto,T. *et al.* (2011) Comprehensive Sequence Analysis of 24,783 Barley Full-Length cDNAs Derived from 12 Clone Libraries. *Plant Physiol.* 156, 20-28

76 Tonooka,Y. and Fujishima,M. (2009) Comparison and critical evaluation of PCR-mediated methods to walk along the sequence of genomic DNA. *Applied Microbiology and Biotechnology* 85, 37-43

77 Laurie,D.A. *et al.* (1995) RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (*Hordeum vulgare* L.) cross. *Genome* 38, 575-585

78 Komatsuda,T. *et al.* (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proceedings of the National Academy of Sciences of the United States of America* 104, 1424-1429

79 Houston,K. *et al.* (2012) Analysis of the barley bract suppression gene *Trd1*. *Theoretical and Applied Genetics* 125, 33-45

80 Rostoks,N. *et al.* (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proceedings of the National Academy of Sciences of the United States of America* 103, 18656-18661

81 Price,A.L. *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38, 904-909

82 Mackay,I. and Powell,W. (2007) Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science* 12, 57-63

83 Meinel,A. (2003) An early scientific approach to heredity by the plant breeder Wilhelm Rimpau (1842-1903). *Plant Breeding* 122, 195-198

84 Harlan,H.V. and Hayes,H.K. (1920) Occurence of the Fixed Intermediate, *Hordeum Intermedium Haxtoni*, in Crosses Between *H. Vulgare Pallidum* and *H.Distichon Palmella*. *Journal of Agricultural Research* 19, 575-599

85 Robertson,D.W. *et al.* (1941) A Summary of Linkage Studies in Barley. *Journal of American Society of Agronomy* 33, 47-64

86 Gustafsson,Å. (1947) Mutations in Agricultural Plants. *Hereditas* 33, 1-100

87 Leonard,W.H. (1942) INHERITANCE OF FERTILITY IN THE LATERAL SPIKELETS OF BARLEY. *Genetics* 27, 299-316

88 Woodward,R.W. (1949) The Inheritance of Fertility in the Lateral Florets of the Four Barley Groups. *Agronomy Journal* 41, 317-322

89 Persson,G. (1969) An Attempt to Find Suitable Genetic Markers for Dense Ear Loci in Barley 1. *Hereditas-Genetiskt Arkiv* 62, 25-96

90 Nybom,N. (1954) Mutation Types in Barley. *Acta Agriculturae Scandinavica* 4, 430-455

91 Gustafsson,Å. and Ekman,G. (1967) Yield efficiency of the X-ray mutant svalöf's pallas barley. *Der Züchter \ Genetics and Breeding Research* 37, 42-46

92 Lundqvist,U. and Lundqvist,A. (1988) Mutagen specificity in barley for 1580 *eceriferum* mutants localized to 79 loci. *Hereditas* 108, 1-12

93 Nordic Genetic Resource Center (2015) http://www.nordgen.org/index.php/en/content/view/full/344

94 Lundqvist,U. and Lundqvist,A. (1988) Induced *intermedium* Mutants in Barley - Origin, Morphology and Inheritance. *Hereditas* 108, 13-26

95 Lundqvist,U. and Franckowiak,J.D. (1997) intermedium spike-f. *Barley Genetics Newsletter* 26, 469

96 Lundqvist,U. and Franckowiak,J.D. (1997) intermedium spike-b. *Barley Genetics Newsletter* 26, 268

97 Lundqvist,U. and Franckowiak,J.D. (1997) intermedium spike-h. *Barley Genetics Newsletter* 26, 470

98 Lundqvist,U. and Franckowiak,J.D. (1997) intermedium spike-i. *Barley Genetics Newsletter* 26, 471

99 Lundqvist,U. and Franckowiak,J.D. (1997) vrs4. *Barley Genetics Newsletter* 26, 159-160

100 Lundqvist,U. and Franckowiak,J.D. (1997) vrs2. *Barley Genetics Newsletter* 26, 263

101 Franckowiak,J. and Lundqvist,U. (1997) Low Number of tillers 1. *Barley Genetics Newsletter* 26, 153

102 Lundqvist,U. and Franckowiak,J. (2007) vrs1. *Barley Genetics Newsletter* 37, 192-194

103 Lundqvist,U. and Franckowiak,J. (2007) intermedium spike-c. *Barley Genetics Newsletter* 37, 237-239

104 Lundqvist,U. and Franckowiak,J. (2007) intermedium spike-k. *Barley Genetics Newsletter* 37, 279

105 Lundqvist,U. and Franckowiak,J. (2007) intermedium spike-m. *Barley Genetics Newsletter* 37, 280

106 Fukuyama,T. *et al.* (1975) Genetic and linkage studies of the five types of induced 'six-row' mutants. *Barley Genetics Newsletter* 5, 12

107 Fukuyama,T. *et al.* (1982) Genetic studies on the induced six-rowed mutants in barley. *Berichte des Ohara Instituts für Landwirtschaftliche Biologie* 18, 99-113

108 Lundqvist,U. (1990) Coordinator's report: ear morphology genes. *Barley Genetics Newsletter* 20, 85-86

109 Gymer,P.T. (1977) Probable allelism of *Ii* and *Int-c* genes. *Barley Genetics Newsletter* 7, 34-35

110 Abebe,B. *et al.* (1990) Enhancing Interaction of Recessive *intermedium* Genes of Barley in Combinations with the 6-Row Gene *hex-v*. *Hereditas* 112, 21-32

111 Abebe,B. (1990) Lateral Floret Development in Barley As Influenced by Recessive Int Genes in a Heterozygous State II. Interaction with Semidominant *int-d* Genes. *Hereditas* 112, 39-48

112 GUSTAFSSON,Å. and Lundqvist,U. (1980) *Hexastichon* and *intermedium* mutants in barley. *Hereditas* 92, 229-236

113 Lundqvist,U. (1996) Barley Genetics Stock Descriptions. *Barley Genetics Newsletter* 26

114 Druka,A. *et al.* (2011) Genetic dissection of barley morphology and development. *Plant Physiol.* 155, 617-627

115 Franckowiak,J.D. *et al.* (1985) Registration of Bowman Barley. *Crop Science* 25, 883

116 Persson,G. (1969) An Attempt to Find Suitable Genetic Markers for Dense Ear Loci in Barley.2. *Hereditas-Genetiskt Arkiv* 63, 1-28

117 He,C.F. *et al.* (2004) AFLP targeting of the 1-cM region conferring the vrs1 gene for six-rowed spike in barley, *Hordeum vulgare* L. *Genome* 47, 1122-1129

118 Harris,J.C. *et al.* (2011) Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. *New Phytol* 190, 823-837

119 Cuesta-Marcos,A. *et al.* (2010) Genome-wide SNPs and re-sequencing of growth habit and inflorescence genes in barley: implications for association mapping in germplasm arrays varying in size and structure. *BMC Genomics* 11, 707

120 Sakuma,S. *et al.* (2010) Duplication of a well-conserved homeodomain-leucine zipper transcription factor gene in barley generates a copy with more specific functions. *Functional & Integrative Genomics* 10, 123-133

121 Sakuma,S. *et al.* (2013) Divergence of expression pattern contributed to neofunctionalization of duplicated HD-Zip I transcription factor in barley. *New Phytol* 197, 939-948

122 Pourkheirandish,M. *et al.* (2007) Analysis of the barley chromosome 2 region containing the six-rowed spike gene vrs1 reveals a breakdown of rice-barley micro collinearity by a transposition. *Theoretical and Applied Genetics* 114, 1357-1365

123 Breja,P. (2014) Functional validation of rice homeodomain genes (*OsHOX14, OsBLH*) and F-Box gene (*OsFBLD10*) in reproductive development and *OsHOX22* in abiotic stress tolerance, University of Delhi

124 Martìn-Trillo,M. and Cubas,P. (2010) TCP genes: a family snapshot ten years later. *Trends in Plant Science* 15, 31-39

125 Doebley,J. *et al.* (1997) The evolution of apical dominance in maize. *Nature* 386, 485-488

126 Clark,R.M. *et al.* (2006) A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nature Genetics* 38, 594-597

127 Takeda,T. *et al.* (2003) The *OsTB1* gene negatively regulates lateral branching in rice. *The Plant Journal* 33, 513-520

128 Lewis,J.M. *et al.* (2008) Overexpression of the maize *Teosinte Branched1* gene in wheat suppresses tiller development. *Plant Cell Reports* 27, 1217-1225

129 Dabbert,T. *et al.* (2010) The genetics of barley low-tillering mutants: *low number of tillers-1* (*lnt1*). *Theoretical and Applied Genetics* 121, 705-715

130 Koppolu,R. et al. (2013) Six-rowed spike4 (Vrs4) controls spikelet determinacy and row-type in barley. Proceedings of the National Academy of Sciences of the United States of America 110, 13198-13203

131 Bortiri,E. *et al.* (2006) *ramosa2* Encodes a LATERAL ORGAN BOUNDARY Domain Protein That Determines the Fate of Stem Cells in Branch Meristems of Maize. *The Plant Cell* 18, 574-585

132 Lundqvist,U. and Lundqvist,A. (1988) Gene Interaction of Induced *intermedium* Mutations of 2-Row Barley I. Double Mutant Recombinants. *Hereditas* 108, 133-140

133 Lundqvist,U. *et al.* (1989) Gene Interaction of Induced *intermedium* Mutations of 2-Row Barley V. Triple Gene Constellations of the *hex-v* Gene and Recessive *int* Genes. *Hereditas* 111, 37-47

134 Lundqvist,U. and Lundqvist,A. (1990) Progressive Promotion of Lateral Floret Development in 3- and 4-Gene Combinations of *intermedium* Genes in Barley. *Hereditas* 113, 237-242

135 Fukuyama,T. *et al.* (1972) A test for allelism of 32 induced six-rowed mutants**.** *Barley Genetics Newsletter* 2, 25-27

136 Lundqvist,U. and Franckowiak,J.D. (2012) Barley Genetics Stock Description 315 (*vrs3*). *Barley Genetics Newsletter* 42, 368-369

137 USDA (2014) PI 383929, Hakata-2, National Small Grains Germplasm Research Facility

138 USDA,A.N.G.R.P.G.R.I.N.G. (2014) PI 189763-Bonus, National Small Grains Germplasm Research Facility

139 Cantalapiedra,C.P. and Contreras-Moreira,B. (2014) Barleymap http://floresta.eead.csic.es/barleymap/

140 Pope,M.X. (1944) SOME NOTES ON TECHNIQUE IN BARLEY BREEDING. *J Hered* 35, 99-111

141 VSN International (2011) GenStat *for Windows* 14th Edition.

142 Moragues,M. *et al.* (2010) Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. *Theoretical and Applied Genetics* 120, 1525-1534

143 Muñoz-Amatriaín,M. *et al.* (2011) An Improved Consensus Linkage Map of Barley Based on Flow-Sorted Chromosomes and Single Nucleotide Polymorphism Markers. *The Plant Genome* 4, 238-249

144 Stam,P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *The Plant Journal* 3, 739-744

145 Haldane,J.B.S. (1919) The combination of linkage values, and the calculation of distances between linked factors. *Journal of Genetics* 8, 299-309

146 Bayer,M. *et al.* (2011) Comparative visualization of genetic and physical maps with Strudel. *Bioinformatics* 27, 1307-1308

147 Ouyang,S. *et al.* (2007) The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Research* 35, D883-D887

148 Klose,R.J. *et al.* (2006) JmjC-domain-containing proteins and histone demethylation. *Nature Reviews Genetics* 7, 715-727

149 Mayer,K.F.X. *et al.* (2009) Gene content and virtual gene order of barley chromosome 1H. *Plant Physiol.* 151, 496-505

150 Baker,K. *et al.* (2014) The low-recombining pericentromeric region of barley restricts gene diversity and evolution but not gene expression. *The Plant Journal* 79, 981-992

151 Yagil,E. and Stebbins,G.L. (1969) The Morphogenetic Effects of the Hooded Gene in Barley II. Cytological and Environmental Factors Affecting Gene Expression. *Genetics* 62, 307-319

152 Boden,S.A. *et al.* (2015) *Ppd-1* is a key regulator of inflorescence architecture and paired spikelet development in wheat. *Nature Plants* 1, 14016

153 Turner,A. *et al.* (2005) The Pseudo-Response Regulator *Ppd-H1* Provides Adaptation to Photoperiod in Barley. *Science* 310, 1031-1034

154 Counce,P.A. *et al.* (1996) Panicle emergence of tiller types and grain yield of tiller order for direct-seeded rice cultivars. *Field Crops Research* 47, 235-242

155 Hussien,A. *et al.* (2014) Genetics of Tillering in Rice and Barley. *The Plant Genome* 7

156 Kirby,E.J.M. and Riggs,T.J. (1978) Developmental Consequences of 2-Row and 6-Row Ear Type in Spring Barley; 2. Shoot Apex, Leaf and Tiller Development. *Journal of Agricultural Science* 91, 207-216

157 Garcia del Moral,M.B. and Garcia del Moral,L.F. (1995) Tiller production and survival in relation to grain yield in winter and spring barley. *Field Crops Research* 44, 85-93

158 Dabbert,T. *et al.* (2009) The genetics of barley low-tillering mutants: *absent lower laterals* (*als*). *Theoretical and Applied Genetics* 118, 1351-1360

159 Faure,S. *et al.* (2012) Mutation at the circadian clock gene *EARLY MATURITY 8* adapts domesticated barley (*Hordeum vulgare*) to short growing seasons. *Proceedings of the National Academy of Sciences of the United States of America* 109, 8328-8333

160 Borràs-Gelonch,G. *et al.* (2012) Genetic control of pre-heading phases in the Steptoe x Morex barley population under different conditions of photoperiod and temperature. *Euphytica* 183, 303-321

161 Malosetti,M. *et al.* (2011) Gene and QTL detection in a three-way barley cross under selection by a mixed model with kinship information using SNPs. *Theor Appl Genet* 122, 1605-1616

162 Thomas,W.T.B. *et al.* (1991) The effects of major genes on quantitatively varying characters in barley. 4. The *GPert* and *denso* loci and quality characters. *Heredity* 66, 381-389

163 Powell,W. *et al.* (1985) The effects of major genes on quantitatively varying characters in barley 2. The *denso* and daylength response loci. *Heredity* 54, 349-352

164 Alqudah,A.M. and Schnurbusch,T. (2014) Awn primordium to tipping is the most decisive developmental phase for spikelet survival in barley. *Funct. Plant Biol.* 41, 424-436

165 Cottrell,J.E. *et al.* (1985) A comparison of spike and spikelet survival in mainstem and tillers of barley. *Annals of Applied Biology* 106, 365-377

166 Arisnabarreta,S. and Miralles,D.J. (2004) The influence of fertiliser nitrogen application on development and number of reproductive primordia in field-grown two- and six-rowed barleys. *Aust. J. Agric. Res.* 55, 357-366

167 Kirby,E.J.M. and Rymer,J.L. (1974) Development of the Vascular System in the Ear of Barley. *Annals of Botany* 38, 565-573

168 Miralles,D.J. *et al.* (2000) Duration of the stem elongation period influences the number of fertile florets in wheat and barley. *Funct. Plant Biol.* 27, 931-940

169 Mascher,M. *et al.* (2014) Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biology* 15, R78

170 Mascher,M. *et al.* (2013) Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *The Plant Journal* 76, 494-505

171 González-Neira,A. (2013) The GoldenGate Genotyping Assay: Custom Design, Processing, and Data Analysis. In *Pharmacogenomics* (Innocenti,F. and van Schaik,R.H.N., eds), pp. 147-153, Humana Press

172 Briggs,S. (1992) A suppressor of floral leaf development. *Maize Genetics Cooperative Newsletter* 66, 50-51

173 Nagao,S. and Takahashi,M. (1963) Trial Construction of Twelve Linkage Groups in Japanese Rice: (Genetical Studies on Rice Plant, XXVII). Journal of the Faculty of Agriculture, Hokkaido University 53, 72-130

174 Franckowiak,J. and Lundqvist,U. (1997) BGS 202; Third outer glume. *Barley Genetics Newsletter* 26, 207-209

175 Wang,L. *et al.* (2009) *NECK LEAF 1*, a GATA type transcription factor, modulates organogenesis by regulating the expression of multiple regulatory genes during reproductive development in rice. *Cell Research* 19, 598-611

176 Mackill,D.J. *et al.* (1991) Frizzy panicle, an EMS-induced mutant in the Japonica cultivar M-201. *Rice Genetics Newsletter* 9, 100-102

177 Kempton,J. (1934) HERITABLE CHARACTERS IN MAIZE: XLVII. Branched Silkless. *J Hered* 25, 29-32

178 Franckowiak,J.D. (1997) Barley Mutant Stock Description-Compositum 2. *Barley Genetics Newsletter* 26, 71

179 Chuck,G. *et al.* (2002) The Control of Spikelet Meristem Identity by the *branched silkless1* Gene in Maize. *Science* 298, 1238-1241

180 Komatsu,M. *et al.* (2003) *FRIZZY PANICLE* is required to prevent the formation of axillary meristems and to establish floral meristem identity in rice spikelets. *Development* 130, 3841-3850

181 Poursarebani,N. *et al.* (2015) The Genetic Basis of Composite Spike Form in Barley and "Miracle-Wheat". *Genetics*

182 EHRENBERG,L. *et al.* (1959) THE MUTAGENIC EFFECTS OF IONIZING RADIATIONS AND REACTIVE ETHYLENE DERIVATIVES IN BARLEY. *Hereditas* 45, 351-368

183 USDA ARS National Genetic Resources Program (2014) *Germplasm Resources Information Network - (GRIN)*

184 Cockram,J. *et al.* (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proceedings of the National Academy of Sciences of the United States of America* 107, 21611-21616

185 Bockelman,H.E. and Valkoun,J. (2011) Barley germplasm conservation and resources. In *Barley: production, improvement, and uses* (Ullrich,S.E., ed), Wiley-Blackwell

186 Pasam,R.K. *et al.* (2012) Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biology* 12, 16

187 Benson,D.A. *et al.* (2013) GenBank. *Nucleic Acids Research* 41, D36-D42

188 Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792-1797

189 Mitchell,A. *et al.* (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Research* 43, D213-D221

190 Untergasser,A. *et al.* (2012) Primer3–new capabilities and interfaces. *Nucleic Acids Research* 40, e115

191 Zhang,J. and Madden,T.L. (1997) PowerBLAST: A New Network BLAST Application for Interactive or Automated Sequence Analysis and Annotation. *Genome Research* 7, 649-656

192 Hammer,Ø. *et al.* (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4, 9

193 Tondelli,A. *et al.* (2014) Allelic variation at *Fr-H1*/*Vrn-H1* and *Fr-H2* loci is the main determinant of frost tolerance in spring barley. *Environmental and Experimental Botany* 106, 148-155

194 Dabney,A. and Storey,J. (2003) qvalue: Q-value estimation for false discovery rate control R package version 1.43.0.

195 Goodstein,D.M. *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* 40, D1178-D1186

196 The International Wheat Genome Sequencing Consortium ((2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345, 1251788

197 Tamura,K. *et al.* (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30, 2725-2729

198 Capra,J.A. and Singh,M. (2007) Predicting functionally important residues from sequence conservation. *Bioinformatics* 23, 1875-1882

199 Winkel-Shirley,B. (2001) Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. *Plant Physiol.* 126, 485-493

200 Houston,K. *et al.* (2014) A genome wide association scan for (1,3;1,4)-β–glucan content in the grain of contemporary 2-row Spring and Winter barleys. *BMC Genomics* 15, 907

201 Hrmova,M. *et al.* (2007) A Barley Xyloglucan Xyloglucosyl Transferase Covalently Links Xyloglucan, Cellulosic Substrates, and (1,3;1,4)-β-D-Glucans. *Journal of Biological Chemistry* 282, 12951-12962

202 Sun,Q. and Zhou,D.X. (2008) Rice jmjC domain-containing gene JMJ706 encodes H3K9 demethylase required for floral organ development. *Proceedings of the National Academy of Sciences of the United States of America* 105, 13679-13684

203 Botër,M. *et al.* (2007) Structural and Functional Analysis of SGT1 Reveals That Its Interaction with HSP90 Is Required for the Accumulation of Rx, an R Protein Involved in Plant Immunity. *The Plant Cell Online* 19, 3791-3804

204 Kadota,Y. and Shirasu,K. (2012) The HSP90 complex of plants. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1823, 689-697

205 Hiltscher,H. *et al.* (2014) The *radical induced cell death protein 1* (*RCD1*) supports transcriptional activation of genes for chloroplast antioxidant enzymes. *Frontiers in Plant Science* 5, 475

206 Overmyer,K. *et al.* (2000) Ozone-Sensitive Arabidopsis *rcd1* Mutant Reveals Opposite Roles for Ethylene and Jasmonate Signaling Pathways in Regulating Superoxide-Dependent Cell Death. *The Plant Cell Online* 12, 1849-1862

207 Murata,N. and Tasaka,Y. (1997) Glycerol-3-phosphate acyltransferase in plants. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 1348, 10-16

208 Murata,N. *et al.* (1992) Genetically engineered alteration in the chilling sensitivity of plants. *Nature* 356, 710-713

209 Yamakami,M. *et al.* (2003) Tom1, a VHS Domain-containing Protein, Interacts with Tollip, Ubiquitin, and Clathrin. *Journal of Biological Chemistry* 278, 52865-52872

210 Kim,C.M. *et al.* (2007) OsCSLD1, a Cellulose Synthase-Like D1 Gene, Is Required for Root Hair Morphogenesis in Rice. *Plant Physiol.* 143, 1220-1230

211 Yamaguchi,M. *et al.* (2012) Loss of the Plastid Envelope Protein AtLrgB Causes Spontaneous Chlorotic Cell Death in *Arabidopsis thaliana*. *Plant and Cell Physiology* 53, 125-134

212 Yang,Y. *et al.* (2012) A chloroplast envelope membrane protein containing a putative LrgB domain related to the control of bacterial death and lysis is required for chloroplast development in Arabidopsis thaliana. *New Phytol* 193, 81-95

213 Inatsugi,R. *et al.* (2002) Phosphatidylcholine Biosynthesis at Low Temperature: Differential Expression of CTP:Phosphorylcholine Cytidylyltransferase Isogenes in *Arabidopsis thaliana*. *Plant and Cell Physiology* 43, 1342-1350

214 Inatsugi,R. *et al.* (2009) Isozyme-Specific Modes of Activation of CTP:Phosphorylcholine Cytidylyltransferase in *Arabidopsis thaliana* at Low Temperature. *Plant and Cell Physiology* 50, 1727-1735

215 Gui,J. *et al.* (2011) Functional Characterization of Evolutionarily Divergent 4-Coumarate:Coenzyme A Ligases in Rice. *Plant Physiol.* 157, 574-586

216 Kajita,S. *et al.* (1996) Alterations in the Biosynthesis of Lignin in Transgenic Plants with Chimeric Genes for 4-Coumarate: Coenzyme A Ligase. *Plant and Cell Physiology* 37, 957-965

217 Rai,A. *et al.* (2012) Membrane Topology and Predicted RNA-Binding Function of the Early Responsive to Dehydration (ERD4) Plant Protein. *Plos One* 7, e32658

218 Aluri,S. and B++ttner,M. (2007) Identification and functional expression of the *Arabidopsis thaliana vacuolar glucose transporter 1* and its role in seed germination and flowering. *Proceedings of the National Academy of Sciences* 104, 2537-2542

219 Rasmusson,A.G. *et al.* (1998) Physiological, biochemical and molecular aspects of mitochondrial complex I in plants. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1364, 101-111

220 Marienfeld,J.R. and Newton,K.J. (1994) The Maize *Ncs2* Abnormal Growth Mutant Has a Chimeric *nad4-nad7* Mitochondrial Gene and Is Associated with Reduced Complex I Function. *Genetics* 138, 855-863

221 Kodym,A. and Afza,R. (2003) Physical and Chemical Mutagenesis. In *Plant Functional Genomics* (236 edn) (Grotewold,E., ed), pp. 189-203, Humana Press

222 Mulder,N. and Apweiler,R. (2007) InterPro and InterProScan. In *Comparative Genomics* (Bergman,N.H., ed), pp. 59-70, Humana Press

223 NCBI (2015) NCBI Protein Database

224 Quan,Z. *et al.* (2011) JmjN interacts with JmjC to ensure selective proteolysis of Gis1 by the proteasome. *Microbiology* 157, 2694-2701

225 Chen,Z. *et al.* (2006) Structural Insights into Histone Demethylation by JMJD2 Family Members. *Cell* 125, 691-702

226 Tsukada,Y.i. *et al.* (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439, 811-816

227 Reddy,A.S.N. (2007) Alternative Splicing of Pre-Messenger RNAs in Plants in the Genomic Era. *Annu. Rev. Plant Biol.* 58, 267-294

228 Brown,J.W.S. and Simpson,C.G. (1998) SPLICE SITE SELECTION IN PLANT PRE-mRNA SPLICING. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* 49, 77-95

229 Simpson,C.G. *et al.* (2008) Monitoring changes in alternative precursor messenger RNA splicing in multiple gene transcripts. *The Plant Journal* 53, 1035-1048

230 Valdar,W.S. (2002) Scoring residue conservation. *Proteins: structure, function, and bioinformatics* 48, 227-241

231 Fisk,S. *et al.* (2013) *FR-H3*: a new QTL to assist in the development of fall-sown barley with superior low temperature tolerance. *Theoretical and Applied Genetics* 126, 335-347

232 Yokoo,T. *et al.* (2014) *Se14*, Encoding a JmjC Domain-Containing Protein, Plays Key Roles in Long-Day Suppression of Rice Flowering through the Demethylation of H3K4me3 of *RFT1*. *Plos One* 9, e96064

233 Li,T. *et al.* (2013) Jumonji C Domain Protein JMJ705-Mediated Removal of Histone H3 Lysine 27 Trimethylation Is Involved in Defense-Related Gene Activation in Rice. *The Plant Cell Online* 25, 4725-4736

234 Shen,Y. *et al.* (2014) Over-expression of histone H3K4 demethylase gene JMJ15 enhances salt tolerance in *Arabidopsis*. *Frontiers in Plant Science* 5, 290

235 Lu,F. *et al.* (2011) Arabidopsis *REF6* is a histone H3 lysine 27 demethylase. *Nature Genetics* 43, 715-719

236 Lan,F. *et al.* (2007) A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* 449, 689-694

237 Whetstine,J.R. *et al.* (2006) Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* 125, 467-481

238 Wolffe,A. (1998) *Chromatin: structure and function,* Academic Press

239 Pfluger,J. and Wagner,D. (2007) Histone modifications and dynamic regulation of genome accessibility in plants. *Current Opinion in Plant Biology* 10, 645-652

240 Tsai,C.-L. *et al.* (2014) How substrate specificity is imposed on a histone demethylase lessons from KDM2A. *Genes & Development* 28, 1735-1738

241 Kooistra,S.M. and Helin,K. (2012) Molecular mechanisms and potential functions of histone demethylases. *Nature Reviews Molecular Cell Biology* 13, 297-311

242 Berry,W.L. and Janknecht,R. (2013) KDM4/JMJD2 Histone Demethylases: Epigenetic Regulators in Cancer Cells. *Cancer Research* 73, 2936-2942

243 Hu,Y. and Lai,Y. (2015) Identification and expression analysis of rice histone genes. *Plant Physiology and Biochemistry* 86, 55-65

244 Papaefthimiou,D. and Tsaftaris,A. (2012) Significant induction by drought of *HvPKDM7-1*, a gene encoding a jumonji-like histone demethylase homologue in barley (*H. vulgare*). *Acta Physiologiae Plantarum* 34, 1187-1198

245 Lamesch,P. *et al.* (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40, D1202-D1210

246 Ahokas,H. (1976) A Mutant of Barley: Awned Palea. *Barley Genetics Newsletter* 7, 8-10

247 Lundqvist,U. and Franckowiak,J. (1996) Barley Genetics Stock Description Special Edition (26 edn)

248 Yuo,T. *et al.* (2012) A *SHORT INTERNODES* (*SHI*) family transcription factor gene regulates awn elongation and pistil morphology in barley. *Journal of Experimental Botany* 63, 5223-5232

249 Lundqvist,U. and Franckowiak,J. (2003) Diversity of Barley mutants. In *Diversity in Barley (Hordeum vulgare)* (Von Bothmer,R., ed), pp. 77-96, Elsevier

250 Jost,M. *et al.* (2014) Cloning of the Gene *Laxatum* (*lax.a*) - Prospects from an Improving Barley Genomics Infrastructure. In *Plant and Animal Genome XXII*

251 Casao,M.C. *et al.* (2011) Adaptation of barley to mild winters: A role for PPDH2. *BMC Plant Biology* 11, 164

252 Faure,S.+. *et al.* (2007) The *FLOWERING LOCUS T-Like* Gene Family in Barley (*Hordeum vulgare*). *Genetics* 176, 599-609

253 Deng,W. *et al.* (2015) Direct links between the vernalization response and other key traits of cereal crops. *Nature Communications* 6, 5882

254 Matthies,I.E. *et al.* (2014) Genome-Wide Association Mapping for Kernel and Malting Quality Traits Using Historical European Barley Records. *Plos One* 9, e110046

255 Saisho,D. *et al.* (2009) Allelic variation of row type gene Vrs1 in barley and implication of the functional divergence. *Breeding Science* 59, 621-628

256 Ullrich,S. (2010) Significance, Adaptation, Production and Trade of Barley (Ulrich,S., ed),

257 Thiel,T. *et al.* (2009) Evidence and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice. *BMC Evolutionary Biology* 9, 209

258 Qian,S. *et al.* (2015) Expansion and Functional Divergence of Jumonji C-Containing Histone Demethylases: Significance of Duplications in Ancestral Angiosperms and Vertebrates. *Plant Physiol.* 168, 1321-1337

259 Chaw,S.M. *et al.* (2004) Dating the Monocot Dicot Divergence and the Origin of Core Eudicots Using Whole Chloroplast Genomes. *Journal of Molecular Evolution* 58, 424-441

260 Salse,J. *et al.* (2008) Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution. *The Plant Cell Online* 20, 11-24

261 HGCA (2005) The Barley Growth Guide, HGCA

262 Elliot,W.A. and Poehlman,J.M. (1975) Inheritance of Kernel-Weight in Six-rowed x Two-rowed Barley Crosses (*H. vulgare* L. x *H.distichum* L.). In *3rd International Barley Genetics Symposium* pp. 678-685

263 Gebhardt,D.J. *et al.* (1993) Kernel Morphology and Malting Quality Variation in Lateral and Central Kernels of Six-row Barley. *American Society of Brewing Chemists* 51, 145-148

264 Schwarz,P.B. and Horsley,R.D. (1995) Malting Quality Improvement in North American Six-rowed Barley Cultivars since 1910. *Journal of the American Society of Brewing Chemists* 53, 14-18

265 Gustafsson,Å. and Lundqvist,U. (1980) *Hexastichon* and *intermedium* Mutants in Barley. *Hereditas* 92, 229-236

266 Lundqvist,U. and Lundqvist,A. (1987) An *intermedium* Gene Present in A Commercial 6-Row Variety of Barley. *Hereditas* 107, 131-135

267 Lundqvist,U. *et al.* (1988) Gene Interaction of Induced *intermedium* Mutations of 2-Row Barley .2. Interaction Between the *hex-v* Gene and *int* Genes. *Hereditas* 109, 197-204

268 Lundqvist,U. *et al.* (1988) Gene Interaction of Induced *intermedium* Mutations of 2-Row Barley .3. Overlapping in Dihybrid F_2 Classification Patterns in Combinations of Recessive *int* Genes. *Hereditas* 109, 205-214

269 Lundqvist,U. *et al.* (1989) Gene Interaction of Induced *intermedium* Mutations of 2-Row Barley .4. Overlapping in Dihybrid F_2 Classification Patterns in Combinations of *int-d* and Recessive *int* Genes. *Hereditas* 110, 117-126

270 Lundqvist,U. and Lundqvist,A. (1991) Enhancing Interaction Between Intermedium Genes of Barley in 3-Gene Or 4-Gene Segregations. *Cereal Research Communications* 19, 177-194

271 Ayoub,M. *et al.* (2002) QTLs affecting kernel size and shape in a two-rowed by six-rowed barley cross. *Theoretical and Applied Genetics* 105, 237-247

272 Rasmusson,D. and Wilcoxson,R. (1979) Registration of Morex barley. (Reg. No. 158). *Crop Science* 19, 293

273 Mather,K. (1938) *The Measurement of Linkage in Heredity*, Methuen and Co. Ltd

274 Li,H. *et al.* (2010) Construction of a high-density composite map and comparative mapping of segregation distortion regions in barley. *Molecular Genetics and Genomics* 284, 319-331

275 Sayed,H. *et al.* (2002) Segregation distortion in doubled haploid lines of barley (*Hordeum vulgare* L.) detected by simple sequence repeat (SSR) markers. *Euphytica* 125, 265-272

276 Gymer,P. (1978) The genetics of the six-row/two-row character. *Barley Genetics Newsletter* 8, 44-46

277 Lundqvist,U. (1996) Globosum Induced Mutants. *Barley Genetics Newsletter* 26

278 Luo,J. *et al.* (2013) *An-1* Encodes a Basic Helix-Loop-Helix Protein That Regulates Awn Development, Grain Size, and Grain Number in Rice. *Plant Cell* 25, 3360-3376

279 Gu,B. *et al.* (2015) *An-2* Encodes a Cytokinin Synthesis Enzyme That Regulates Awn Length and Grain Production in Rice. *Molecular Plant*

280 Cottrell,J.E. and Dale,J.E. (1984) VARIATION IN SIZE AND DEVELOPMENT OF SPIKELETS WITHIN THE EAR OF BARLEY. *New Phytol* 97, 565-573

281 Hubbard,L. *et al.* (2002) Expression Patterns and Mutant Phenotype of *teosinte branched1* Correlate With Growth Suppression in Maize and Teosinte. *Genetics* 162, 1927-1935

282 Finlayson,S.A. (2007) *Arabidopsis TEOSINTE BRANCHED1-LIKE 1* Regulates Axillary Bud Outgrowth and is Homologous to Monocot *TEOSINTE BRANCHED1*. *Plant and Cell Physiology* 48, 667-677

283 Lukens,L.N. and Doebley,J. (1999) Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. *Genetical Research* 74, 291-302

284 Kebrom,T.H. *et al.* (2010) Suppression of sorghum axillary bud outgrowth by shade, phyB and defoliation signalling pathways. *Plant, Cell & Environment* 33, 48-58

285 Kebrom,T.H. *et al.* (2006) *Phytochrome B* Represses *Teosinte Branched1* Expression and Induces Sorghum Axillary Bud Outgrowth in Response to Light Signals. *Plant Physiol.* 140, 1109-1117

286 Xu,C. *et al.* (2015) Differential expression of *GS5* regulates grain size in rice. *Journal of Experimental Botany* 66, 2611-2623

287 Whipple,C.J. *et al.* (2011) *grassy tillers1* promotes apical dominance in maize and responds to shade signals in the grasses. *Proceedings of the National Academy of Sciences of the United States of America* 108, 13375-13376

288 Lekanne Deprez,R.H. *et al.* (2002) Sensitivity and accuracy of quantitative real-time polymerase chain reaction using SYBR green I depends on cDNA synthesis conditions. *Analytical Biochemistry* 307, 63-69

289 Valasek,M.A. and Repa,J.J. (2005) The power of real-time PCR. *Advances in Physiology Education* 29, 151-159

290 Rensink,W.A. and Buell,C.R. (2005) Microarray expression profiling resources for plant genomics. *Trends in Plant Science* 10, 603-609

291 Druka,A. *et al.* (2006) An atlas of gene expression from seed to seed through barley development. *Functional & Integrative Genomics* 6, 202-211

292 Guo,P. *et al.* (2009) Differentially expressed genes between drought-tolerant and drought-sensitive barley genotypes in response to drought stress during the reproductive stage. *Journal of Experimental Botany* 60, 3531-3544

293 Wang,Z. *et al.* (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57-63

294 Sims,D. *et al.* (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics* 15, 121-132

295 Tarazona,S. *et al.* (2011) Differential expression in RNA-seq: A matter of depth. *Genome Research* 21, 2213-2223

296 de Almeida Engler,J. *et al.* (2001) In Situ Hybridization to mRNA of Arabidopsis Tissue Sections. *Methods* 23, 325-334

297 Sekhon,R.S. *et al.* (2011) Genome-wide atlas of transcription during maize development. *The Plant Journal* 66, 553-563

298 Komatsuda,T. *et al.* (1997) Identification of random amplified polymorphic DNA (RAPD) markers linked to the *v* locus in barley, *Hordeum vulgare* L. *Theoretical and Applied Genetics* 95, 637-642

299 Wang,L. *et al.* (2010) A dynamic gene expression atlas covering the entire life cycle of rice. *The Plant Journal* 61, 752-766

300 The James Hutton Institute (2014) barleyGenes - Barley RNA-seq Database

301 Czechowski,T. *et al.* (2005) Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis. *Plant Physiol.* 139, 5-17

302 Pfaffl,M.W. *et al.* (2002) Relative expression software tool (REST-©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research* 30, e36

303 Vandesompele,J. *et al.* (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 3, research0034

304 Minakuchi,K. *et al.* (2010) *FINE CULM1* (*FC1*) Works Downstream of Strigolactones to Inhibit the Outgrowth of Axillary Buds in Rice. *Plant and Cell Physiology* 51, 1127-1135

305 Gaudin,A.C.M. *et al.* (2014) The effect of altered dosage of a mutant allele of *Teosinte branched 1* (*tb1-ref*) on the root system of modern maize. *BMC Genetics* 15, 23

306 Bengough,A.G. *et al.* (2004) Gel observation chamber for rapid screening of root traits in cereal seedlings. *Plant and Soil* 262, 63-70

307 Downie,H. *et al.* (2012) Transparent Soil for Imaging the Rhizosphere. *Plos One* 7, e44276

308 Bai,F. *et al.* (2012) TCP transcription factor, *BRANCH ANGLE DEFECTIVE 1* (*BAD1*), is required for normal tassel branch angle formation in maize. *Proceedings of the National Academy of Sciences of the United States of America* 109, 12225-12230

309 Liu,H. *et al.* (2014) OsmiR396d-Regulated OsGRFs Function in Floral Organogenesis in Rice through Binding to Their Targets *OsJMJ706* and *OsCR4*. *Plant Physiol.* 165, 160-174

310 Higo,K. *et al.* (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research* 27, 297-300

311 Lundqvist,U. (1981) Intermedium and hexastichon mutants in barley. Barley genetics IV. Proceedings of the fourth international barley genetics symposium, Edinburgh, 22-29 July 1981

8. Appendices

Appendix 1. Fine mapping KASP assay design nucleotide sequences

(Differential SNP are highlighted in red and surrounded by square brackets)

SNP ID (cM on 1H)[a]	Nucleotide sequence submitted for KASP assay design
12_30562 (42.42)	AACGTAGTCAAAGTATAACATAAATAACTTCAGCAACTGAATGGCAATACATTCC TTGTGCTACGGCATCGTCCACCTATTACTATAAGTTGACCAGGACGTCTTATTTA TATTTATTTTGGTAGTGCTAACTGCGCTGCTCATGTTATCCAATCTTGCC[A/G] CTCAGCCGGCATGTTCCTGCTTCAGACATGCTTCGATCCTGTAACAATAAACCCA AAACACAATCAGCTCAATCTGTTACTCCCTC
12_31208 (42.42)	GCATTGTTTGCTGGTAAAGATGCCAGTAGGGCATTGGCGAAGATGTCCTTTGAGC CACAAGATCTGACAGGTGATGTATCTGGCCTTGGTTCATTTGAGCTTAGTGCCTT GCAGGACTGGGAGTACAAGTTCACCAGCAAGTATGTGAAAGTAGGAAGCATCAAA GGAACAGGACCCATAGAAGAGGGTAATGTTAGCACTACCTCTGAAATA[C/G]AG GAGGAAGCTGCTGCGGTGAAACTGAATGGAGAGAAAGCACCATGAGACACTACTG GAAATATAATTGACTCTGAATTTCGACGAGAAAGAAAGAAATGGAGGTCACTTCT GATCAAGGATCCAAAGAGAGTCAGAAGCTGGATGCCTCCTAGCTGCTACAGTGCT GTTACCTACATGCATCTGGAGTGTAATTCACAGATATACTGTATAGGTTTGTCTG ACCAGAGGTTGAAGTGTAATGTTAAGCATGGAATATCGAATGAATAGTATGATGC TTCATTGTTGCACTCCACAAGATAATGTTATCGATTTTCCTTTGTC
11_21357 (44.59)	CCGCAGTTCGCAGGACTGTTTCAAGGCAGCATGGGCCTCTCACAAGTCTGTACAC ACGAAAGTAGA[C/T]GCATTGTCATCTCAGCTGTCTCAAGAAGGCTGGAAATAT TGTTTGAAAAAAGGGCGGACCCGGACACTGGAGCTTCCTCGTTTTGATTGGACAG GGTAAATTTTTTTACCACATATTTATATCTACGTTGTTGTTTTATGCTGATGCTT CTATTTTTTTTATAGCCAACAAGATAAAGTATCTTGCCAGTGCATTTTAGTTTTG TGCTGTACCTCTAGGGCAAACTAC
12_11169 (45.20)	CTTACCAGTGCTATCAAAACGTAACTTAGATGTAGTCCTTCCAACTGATGGCACG AGAAGTATGTGCCTACTCTGCCAGTGGAAGATATTGACTCCTCACACAGATGTAA TTATATTCATCTGTGCAATATCTTTCAATGTAATTATATTCTAACACAGCAAAGA TGCATGATATCTTGTTCCTAGAACGGTTATCTGACCTATGTGTACCCATTATGCT CATGCAGGTTCCAGTCCGTGGCATACCTTCAGCAAATCTGGTGGTTTGAAGTGGT TGGAGAGCTTGATTTCTGCTTCCCTGTGGGAACCTACAGTTTGTACTTCAGGGTT CATCTCGGGAAGTTCAGCAAGCGGTTTGGACGCCGTGTTTGTAGCACTGAGCACG TCCATGGTTGGAACAAGAAGCCGGTGCGCTTCCAGCTCTCCACCTCGGACGGCCA GAATGCATTATCTCAATGCTACCTGGACGAGCCCGGGAGTTGGGTTCTGTACCAT GCAGGCGATTTTGTTTCCTCGAAGCCAGGCCAATCACTTAAGCTGAAGTTCTCCA TGGCGCAGATCGACTGCACGCACACGAAAGGTGGCCTGTG[C/T]GTCGACTCGG TGCTCGTGTACCCCAAGGGCTTTCAGCAGGAGAAGGTGGTGACC
11_21000 (45.85)	TAATAAAATGTTCGGGGTAAAAGAAGAGGACCTTGGAGGATATGCTTCGCAGCAA CAGCTTTCCAGGCACTG[A/C]ACCACTAAATCGGGTCTTGGATCCAGAAAGTAG GCCGCCTGATGCATACGAAAACGTGTGTAAGAGATAACGCGCCAAACATGCTTTA ACACAAAATCGT
11_21312 (45.20)	AAGGGATATCTGCTCTTAAGGAAACATTGCTTGCGAAAAAACAATGAAGAAAGCA AGCGATCAGTCAATCACAAACAGAGCATTAGTTCTGGCAAAGTCAAAGAATCTGT TACTTCTGTTGGATCTGAAGAACAACCTGAAAAAGTAAGCGTGGAAGATCATATC ACGACCGAGGAAGAGAATTCTTGTGTTAGTTCAGAGCAAGAAATGCAGGATAAGA CAACTAAAGCAATTCTAGAGAACCAGGCGGATGAAAACC[C/T]CTTTTCATCTA ATTCACGTGATGACAATGAGAGTGAGAAAAGTGATGTGACCCATGTGGAATCCCA TGATGGTGTGGCCTCAGTGCACCAGTTTTCAGTTGAGGATGAAGAAGAAAGTATG CCCCTGCCCAAATATTTGAACACAGGAGGAAGTATAGAAGCTGAAACTAAGTTAT CTAAGGATGCCCGTCCTGTGAAGTCTGCACGGACAGTTGAGGATTTTGAGACATG GCAAAGGATAATTCGTTTGGATGCAGTTCGTGCTAACAATGAATGGGTTTCCTAC TCTCCATCCCAAGCCGCTGTTACCAGGGAGAAGGCAATTGAATCTGCTTCAGCTG TTTGTCTCAAAGACTATGAGCACT
12_30350 (45.20)	AAGACACTCAATGCAATGTTACAGTTACCCATGGCAAAAGTATGTGAAACTTGGA TCTGTGCTTAGGCACTTTGCATA[C/T]ACTGTTGCTGCACTTCATGGATGCCTC GAATCAGAAATTCAGGTACATCATAC

12_30592 (45.20)	GGTGCTTTCAAACATTGATGGATTCCTCCCCTTGACATATATATATACTTTGGGT TCCCGCAAAAAAGAATTCTGGAAAAAATGGTCAAGTCAGTGTCATATTTGGGGGG AGGATGCAGATAACAACAGTGCACACGGCTAGCCATGAGGCACCAACAAGCCATC CATCAGGCACATGTCTTCTTGACAAAAAGAACATATATAGAACAAAAGGAATTTC CACGAGCTCTTCACTGATCAACGTCTTTGGAGGCGAAAGAACATTATAATGCACA TTGTACAACGCCAAGGACATGTATGGCTCAATGAACTTCTAGAAGACCTGGGCGA CTTTGCTCGCCAGTACACGGTCCCGCTTTTCGATGAACTCG[G/T]AAGGGAACC AACCAGCTTTCCCCCTGCACTCACCTTCAGCCCATCCATTGTTTGACACCT
12_31134 (45.20)	AGGCCACAGAGATGACAGCAGAAGATGAGCCACTAAACCCTTAAAACACACCAGC GGCCATTGACGTGAACCGGGCATGGAACAGTCTCCAACTTCATCTGCTTGTCTGC AGGGCTGCTGAATAAGACGTGAACCCAGTTTAGTTTGCTTTGTTAGGTGCGCTCA CGGTATTCAGTTTTGTCAG[A/G]AGAGAGGACACGGCTCGGCATGTATTGAGAT ATATACACTGCACAGTAGAGCTTTGGCATAGTACCAGATCTGTATTTTCCCTCGT ATTCTTGCATTTGCTAATAGGATGATCGTACTTGTATTTAAATCCAACCTCTTGA
12_31272 (45.20)	TGCATTGTCCACAACAGAAATCATCAGCAGAAGAGAGGTTTGGAACTATTCAAAC AAGAGCCAGCTTGAAGCCACTGAAACTTCAAAACTCATACTGGAGAAGAGCCTGC AGGGGCTTGACATCGCGTTGCTGCTGCGGTAGGACAAGTCTACTCCAGTACCTGC CACTGCGACCCGAAGTAAACCTCTGTGGATCTAGTTGAGTATCCACCATCTGTGA TCCAAAGCAAGATCTGTGAAGATCTCAAGAGATTTCAGTGTCCC[T/C]AATCCA ACAAGCTTGTTCTCAACATAAACTTAATATTGCCGTGCAATTTCCTTTCCCGTTC TAAAAGAGATTAGGCCATGATTGTACATTGTAAGACAATTATATCAGGAGTAAAG GGAACTTTTCTTCCGCGTC
11_21217 (48.29)	CTTGTGAAGAGGCACGCGGCAGAGATATTCACGTCGATCGCCATTGCGTCGACAT TCTCCCTGTACTCGACGGCCATCCT[A/G]GGACGCCTGGTTGGGCTGGAGCCAA CGTTGACCATCTCGATCTTGCCAAGGTGCATAACCGTGGCGTTGGCTCTGAGCAT AGTGTCTTTCTTCGA
12_30672 (unmapped)	CTGGGCTCACGCAGCTGCAGCGGATGGGCGTCGGCCGGTTCCTAGCGGTCCCGGC GCTGGCCGCGGCGGCACTGGTTGAGATGTGGAGGCTGCGCAGCGTCGGGGCCGGC CACAACCTGAGCATAGCGTGGCAGCTCCCCCAGTTCATGCTCATCGCCTGCTCTG ACGTGTTCTGCGGCATAGCCCAGCTCGAGTTCTTCTACTCGGAAGCCCCCATGTC GATGCGGAGCCTATGCTCGGCATTCTCGTTCCTGGCGATGTCGCTTGGTTACTAC CTGAACTCCATGATCATCTCGGCCATCGCTGCTCTGTCAAAGAGCGGCGGCGGGA AGGGCTGGCTTCCCGCCGACTTGAACGACGGGCATCTTGACTACTACTTCTGGCT GTGGGCTG[T/G]GATCGGCGCGGTGAACTTCGTTGTGTACACGGCCTTCGCGAA GAATTATACAGTGAAGAAAGTCGAGCTCCGGTGAACACGGTTTGCTGTTGGTCAC AAGCGTACAGTTCTTATCCTCTTTGTTGTGTATGTACTATCTACAATACAGCTTT GGTGTAATTAAATATCGC
12_30786 (47.21)	GTCCAAGGAAGCAAAGCGAAGACTTCGGAAGCCTGCTTTAGTGCATTTAGCTTAT GCAAAAGTTATAAAGAAGCTCAACTTCTGTGGCTAGGAAATGGATGAATGCAGGA TG[C/G]GGATTGGCGCATGAATCTGATCGATGAAATACTGATTGCCAATTGTAA TCTTCTTAGATTTGTCCTTAGATGCATTATCATCAGATTTTTGTGCTCTTATCTG TTCACCAATACTGATGAGCATTTTATCGTATTATCATCTGCTTGTTTTCTACACA ATCTAGTCTTCTAGTTGATAAAGGCTCCCTTGTGTTTTGTTATGCACACGGAGAA GACTCGCCATAGTGATGATGGCCAACCTGCTGCCAGTAAAGAAACTCAATGTTCC TTATCCTAGCTTCAGATTCAGTGAGCTTCACAT
11_10520 (45.20)	ACGACTGAAGTTCTTTGCTGCTGCTTTATCAGCAATGGCCTGCTGCTGCTTAAAT TAACCTTTCTTGTACTAATCACTACCAGATGTTACAATACCTATAGAGCGAGCTT GTTATGTGCT[A/G]TACTGCTCGGCTGATGCTGCTGAGCTTAAGTCCATTCCCT TGATGTGTAATTACATGCCGATCGGGATAACCTTTGCTTCGATCAGGCTGTTTCC ATCCAATAATCAAGCCATG
11_20798 (unmapped)	ATTCAAAGAGGGGAGCTTTCAGCCTCCAGAGATATCGTCTTCAGTAAATCTACCT GAACAGATCGCCATATTCGGGCAGAAGATCAGCCTGGGGCCTGTTAAGCAGCTGC TGGAGCCTTTGCAGCGAGCATTTGCTAGCATTGCAGGGTCCATCTCTGGGCAGCC ACCTTTAAAGGTACCGATTCCCGGTGACAACAAGGCCAAGTCGTGGCTTTTGACA ACCTACCTCGACAAAGACTTGAGGATATCAAAAGGTGATGGGGGTGGCGTGTTTA TTCTTGCCAAAGAGGGGAGCCCTTTACTAGATTAACTGTACATGTGTAATAGGTG TCATTATTCATATGTGCAAGGTTGAAAACCACGCTACAAGCATTCTTCTCCCCCT TTGATACACAATT[T/A]TCTGGTGGTTTTCTAGAGTCTGACAGAATCAACTAAA TGTTGTCAGTATGGAGTCTCCATACGCTTTCTGCTCTTTGGTAACTATAAGTGAC ATTTTGTGTTAT

11_21361 (48.99)	CCAAAACCACCTCACCCGAGAGAAGCGGCGGCGGCGGCATCGGGCGAGATGATTA AGAGACGCTACTTCAGGCAAGACCACGGCGACAATAGCGCCTCCTCGTCCTCCTC CTCCTCGGGGTCTGATTCGGACGGCGACCCCGCGGAAGAAGAGGTCTCGGAGGAG GAAGTAGAGGCGGAGCAAGAAGAAGAAAAGGAAGAA[C/G]GTGATGAAGAAGAA GAAGCGGAGGTGGAGTCGGGGGAAGAACAGGAAAAGGAAGTAGAGCAGCAGGCTG AAGATGAGGGT
12_30110 (45.20)	GCCGGAGCTTGCTGTACATCGTCATCGTCCTGTTGTGTTTTTGCAAGTAGTGTGT TAGCAGTACGAAGCGCCGGATCCATCGTTGTTAACTATCTCGTGACCCTGTGATC GACGACGACGGTGTGTCGGCGTAGTGCTGTTGAACTGAAAACTGTATATAGGACA AAGATGGATTGCCCGTCGCCGTTCTCCTCTACCTCTAGTGTTATTGTTGTTGTGT GTATGCATGCATCGTTTCTC[A/G]TCTTTGCAGAGCGAGCATGGATGATATAAT GATGTCGCTTGCAAAATGAGATGGATCGGGCG
12_30478 (45.20)	CCACGGCAGAGACGGGGTTGTTCCGCTTTCAGAAGGAGCCTTTGCACATATTGAG AAGCCTCTGCCAAAGCCCACTCCTGAATTTAACCCCTTGGCAGCCAAAGCAGCTA CCATCAGTCTCTCAGGATTTGGAGGCTTTTTCAGCTTCGGTGATTTCTCAAACAA GC[A/G]CAATAAGCAGAACTCCAACAAGAAGAATCCCAACAACCTCCCCCAGGT AATTAT
12_30499 (45.20)	GACAAAAGCTCCAGCAGTTGCTTGCTGTGCAAAATACATGCGAGGCTAGCTGGTT AGACCAGTTGTATGTTCCATGCGCTTTCATTTGCTTGGTCAAATGCCAGAACACA ACAGCTACACCATCACACCAAATGCAACGGCAGATAAAGTCTACTGCTAAGGATT ACATCATCTTTGTTCGATGAAAGCTCCAGC[A/G]ACTGCCTGGTTCGAAATACA GATGACTGACAGGCTAGCTTAGTATATGACTGACAGGCTAGCTTAGTATATCTCA TCATCTGGGACATCCGGCGTAAGAAACTTCTGAGGGTCCTTGTACATTTCCTCAT AGTTGTATTTAGGTTTCACCCTCTCTGGAACCGGCTCAAGTGGGATGATGCTGTC TTCGTCAATCACACCATCAATGATCCCGTAATCCACGGCCTCAAGAGGACCCATG TATCGGTCCCTATCGATGTCTTTCTCTACCTGCTCCA
12_30694 (45.20)	GAAAAGGACAATGTATTGGCATGGCATGAAGCCGAAACGAATCGATGCAATCATT TCCATCCCGGATATTCTCTATGGACAGGTGAAACTAACAAAAAGGAGTGCTTTTA CATGTACAAGAAACGCACACGACACCCGACCTCTGTGTAACGGAAAAAAACTGTA CATGCTATGTACTTGCAAGCCACCTGCAGTCACTCTGAAGAGCACCACCAACTCC AAAATGTTCACGACATCCTCAAGAAGCTGAGATCGGCGTGTTTTTCAACTTGTGC GCTACGAATGCCGCCGCCC[A/G]CTGGCAGGATTTACAGGACCAGTCGGTGGAG GAGGAGATGATGATGGTAGACGGACCAGGGCGGGTTGAGGCGAGGGCGCCGCCGT GAGGGCATAGGTGAGCCTTGAAGACGAGCTCTCGATAAGCCCTGCAACTCCGGCA GCGGAAGTTTGTGCATTGGATGGAAGGAACAGGTCATGTTTCTGGTCTTTGTAGG CGCTCCATAC
12_30750 (45.20)	CTTCAATGTTGAGAAGGTTCAGTACAAGAATGTAGT[A/G]TTCACTGTGTGGGA CGTGGGTGGCCAGGAGAAATTGAGGCCCTTGTGGAGGCACTACTTCAACAACACA GATGCTCT
12_30710 (49.65)	TCAGCCCTTGCCTGTTGGAAGACCTACCAAATCAAGGTGATCAAATAGATGATCT GAAAATTCTGTTTGACTGTATCTGCGAAGCTCTAACAGAGATCCAGGAGAGATAT TTTAGACTCTCTTCTTGGTTATCGTTTCTCAAACACGACATAAGGCCACCCCCAA TTGGAGAAGACCTCGTCGCAGAAGTTGACAAATATGTTGATGATTACCTCCGACA TAGTTTTCCGAGTACATTAGATCATACGATCAAAAAGGATTTGGAAGTTCGGACA TGGATGGACATTCGGTCGAAGACTGAAGGGATCGTTGTGGAGATATGGGAGTTTG TGTTAGACGAACTTATCGATGAGGCTGTTTTCGATTTGTGGATTTGAA[G/A]CT CTGATCAGACCCCCCCTGTCAGTGGTGCTTTGTGATGAGTACATACATGTAGGTA TAACTGTAGAGGTTTTCATGCTGACAATAACACCATTTTACGTGGCGGAATAACA CGGTCTGTTTTGTCTGGCCCGTGCGAGTAAAGTAATGTGAGTTTGAACTTGGACG ACGTTTTGCTCCTCGAAGTTTCAACCTGCAAACCTCCCAGCCGTGCAGTTTGTCT CTAGACTGGATGTAGTTTGGGCAGCCTGTCACAGCCATGGCTCTTGGACGGCTAA AATGCACAGCAGTCTG
12_10198 (51.20)	CGGCCACCGCCTCCAGGCGCCCTTCTCCATGCGCATCACCAACGAGTCCGGCAAG ACGCTGGTGGCCGACAAGGTCATCCCGGCCAACTGGGCGCCCAGCACCTTCTACC GCTCCATCGT[C/T]CAGTACAGCTGAAGCTGCTGAACTCATCATCATTTGGAAT CATATGCATACATTGGCTAGATTTGGTATCGTACTATCACGGATCGTTTTGCATG GGTACTACCGAGGATGTTTTTGATTAAGGTGTTGGGTCGTCGGGTGTCATGTAGG CGTCGTCGTGCATGTGTGTCTTGAATACGTCGGTGGGCTATGGAGGAGAGGCACT TTTGCATGTGCTCTCCCGCGCCCACCG

12_30753 (58.59)	GTAACATGGCAAGATTTTATGGAAAAGGATGAGCAATATATTTCTAGTTTGTCTG GTGCCCTATCGATGACATATTATGAACCAGTAAATTGACTGCTACATGAAGAAGC TTTGATGGCTGGCTAGCGTGTACATGGTGGAGTTCAATCAGGCTGTCATTCTTCT GGAGCAATTACCATTCTACACCTGCGAGTGCAAAATTTTGTGCAGATTATCGTTC TACGCCCGCCTGTAGTTGCTGAGTGCAAGATTTTGTGCAGATTATCATGCTATGC CCACCTATATTTGCTGTGCAAAATTGTGTAC[A\G]TATACGACTCTGAATGATG GCTAGCGATCATGAGATCGTCATTGTACTGCCTGAAACCACTGTTGAGTTGAAAT TCTCTGGGTTGGTTGTAGCTGCTGTTCCTTCTGAACAACTGTAAACCCCTTGAAA TGAAGATTAGAAAACGTGGACTGTGAACCTTTTGAAATAAAAAATAGAAAACGTG AAACC
12_30304 (59.29)	AAATTTCAGGGATGCCCACCTCTTTGAAGGAAGAGCCACGGCAGGCAGCTGCCTC CCCAGATTCTAGCGGCGAACCATCATCCTC[C/T]GGGACAATGAAAGAAGACAG GAACGAAATCCCATAAGCTTCATCTGCATCTTGCAAGGAGATGAGAATATCTCAG TAGTCAGGATGATAGGCGCTGATTTGTGATCGTAAAGATGTAAATAAGCTAGCAT GCTTGAGGCGAGGCCTTATTTGGTTTACAGCGCTGAAAAGCACATTTTCCTGGCG GA
12_30821 (54.54)	AAACAATTCTAGAACAAAAAATGTTGAACAGGCAAATAAGCTCAGAACTTCGGAC ATATTAATTTACCATTTGGCCGAAGGTCTCCCTCTAATGTCACTTCTCTGGATTC TGGAGAGAACTCTTGCTTTGAGCCATTGATGGCTGTAAACGCAACAACCACCTCA GTTGGGTGGAGAAGAGTGAAAGTGGGATGAACTCGCAAACAAACCAACCTTCACA GGAGGAACAAACTTCAATTTTCTGAGATACCAATTCACTCTGAATAATCAAAGAC ACCTCTGACTAAAACTTTAACTACAGGTTGGATCGAACTAATTATAAAACTCTCT AGTTAAAATTGCCGAATGGTA[C/G]AACAAAGCATACTCAGGAAGGGGGAAAAC CCAGCAAACCCAAGATATAATAGAATTTGAAAGCATTTAGTTTTGTGGCATCAAT GCAATCTGTGGGGGCATTCACGTGAGTGCTTTGTGTTGTAATTACATGAGTCGCA AAACTGCGCAATTCCTGATCAAGGATTTACAAACAAAACTTTTTTATGAAACTTA GAACTCACCTTGAAAACCCACCAGAGCCAGCTCCAACACTTCTTGCTTCAATGC

[a]Map position from Muñoz-Amatriaín map

Assay Name and Allele genotyped	Nucleotide sequence
INT-C Int-c.a/*int-c*.b (4H)	GCAGCCGAGATCGACGACCAGTCGCCGNCGGAGCTGCTT CTGATGGATCAGGCGCCGGCGCCCAGGCCAGACGGGG[C /T]TGGAAAGGCACAAGGCCTGCACGGCGGTGGAGGCCT CGACAGCGCGGCCGCTAGGAAAGACCGTCACAGCAAGAT ATGCACCGCCGGCGGGATGAGGGACCGGCGGATGCGGCT GTCCCTC
OLAD489 vrs1.a1/Vrs1 .b3 (2H)	CTCCCAGTCTCTCCTTCAGCCTCAGCAGCTGGCCATATG TTTGTCAGTAAGCATGTTCAAGAACGGAGAGGAACTGCA GATGTGCAGCGCCATATGTAAGCCAGTGTGAGTGTATGC GAGCAAGCATACCTCGTTCTCGAGGTGGCATTTGTGGAG GATGGCGGCGTCGTGGGCGTGCTTGAGCCTCGCGAACTC CTCCTCGAGCGTCTTGTTCTTGTGGCGCGCGCGGCGGTT CTGGAACCACACGGCGACCTGCTTGGGGTCCAGCCCGAG CTCGGCGGCCAGATACACCTTGCGGGCTGTCTCCAGCTT GCGGTCCTCCCGGAAGCTCAGCTCCAGAATCTCGGCCTG CTCGTCGGTGAGCCGCCGCTTCTTGGGGTCCCCTCCTCC GTCCATCTCCCCACCGTCACCATCCCCTCCGCCGCACCT CGCCGACCTCCGCCGCCTGCGCCGCGCCCTCTGCTTGCT GGT[A/C]TCCCCCTGCGCCGTGCCTGCACATACAAACA CAGAATGGAAAAAACGTCATCACATGCGTGTCCTGGCCG GTGAAACTGAA

Appendix 2. Test of association between lateral spikelet awn phenotype and row-type at four different positions across the spike

	Laterals B1		Laterals B2		Laterals M		Laterals T	
F2 Population	X^2	P Value	X^2	P Value	X^2	P Value	X^2	P Value
BW419Ba	72.13	p<0.001	178.84	p<0.001	216	p<0.001	188.15	p<0.001
BW419Bo	67.96	p<0.001	134.44	p<0.001	176	p<0.001	159.85	p<0.001
BW902Ba	82.35	p<0.001	124.19	p<0.001	166.33	p<0.001	166.33	p<0.001
BW902Bo	141.37	p<0.001	212	p<0.001	214	p<0.001	212	p<0.001

(5 degrees of freedom (d.f.) in all cases)

Appendix 3. Test of association between lateral spikelet awn phenotype and Parent Cultivar

Spike Position	X^2	P-Value	d.f.
B1	13.71	0.017	5
B2	12.84	0.025	5
M	18.38	< 0.001	5
T	28.4	< 0.001	5

Appendix 4. Test of association between central spikelet awn phenotype and row-type across F_2 populations

	Central spikelet B1			Central spikelet B2		
F_2 Population	X^2	P Value	d.f.	X^2	P Value	d.f.
BW419Ba	70.58	<0.001	5	11.91	0.036	5
BW419Bo	89.98	< 0.001	5	14.92	0.008	5
BW902Ba	63.08	< 0.001	5	11.56	0.026	5
BW902Bo	100.23	< 0.001	5	16.23	0.004	5

	Central spikelet M			Central spikelet T		
F_2 Population	X^2	P Value	d.f.	X^2	P Value	d.f.
BW419Ba	0	1	0	0	1	0
BW419Bo	0	1	0	0	1	0
BW902Ba	0	1	0	1.51	0.567	5
BW902Bo	0	1	0	0.69	0.709	5

Appendix 5. Test of association between central spikelet awn phenotype of two-rowed spikes at B1 position and parent cultivar

Bowman vs Barke χ^2=17.76 (5df, p<0.001).

Appendix 6. Test of association between central spikelet grain-fill and row-type at the B1 spike position across F_2 populations

	Central spikelet B1		
F_2 Population	X^2	p Value	d.f.
BW419Ba	23.63	<0.001	1
BW419Bo	48.05	< 0.001	1
BW902Ba	18.17	< 0.001	1
BW902Bo	40.94	< 0.001	1

Appendix 7. Test of association between spike row-type and spike phenotypes

Phenotype	Spike Position	χ2	p-Value	d.f.
Multi-florous	B1	11.9	<0.001	1
Multi-florous	B2	78.73	<0.001	1
Multi-florous	M	28.75	<0.001	1
Multi-florous	T	15.48	<0.001	1
Awned-Palea	B1	97.92	<0.001	1
Awned-Palea	B2	152.25	<0.001	1
Awned-Palea	M	147.56	<0.001	1
Awned-Palea	T	85.5	<0.001	1
Broad Awn	B1	10.27	0.008	1
Broad Awn	B2	15.48	<0.001	1
Broad Awn	M	3.39	0.124	1
Broad Awn	T	No observations		
Split Awn	B1	11.61	0.003	1
Split Awn	B2	16.16	<0.001	1
Split Awn	M	25.33	<0.001	1
Split Awn	T	3.9	0.085	1
Multiple Lemma	B1	35.03	0.003	1
Multiple Lemma	B2	41.08	<0.001	1
Multiple Lemma	M	10.27	0.008	1
Multiple Lemma	T	15.48	<0.001	1

Appendix 8. Mean spike emergence(days) across BW419Bo, BW419Ba, BW902Bo and BW902Ba F_2 individuals

Sowing Date		BW419	BW902	
26/03/10	Barke	70.90	74.63	
	Bowman	67.22	66.09	
22/04/10	Barke	55.95	58.25	
	Bowman	51.70	53.26	SED=0.884

Appendix 9. Mean tiller number for the Bowman and Barke F_2 populations (p=0.035)

SED=0.4614 between transformed tiller values. The numbers shown in brackets are the non-transformed equivalent values.

Resown	Parent	Mutant-allele	Row-type	
			Two-rowed	*vrs3* mutant
No	Barke	BW419	7.459 (56)	6.509 (42)
		BW902	6.899 (48)	6.032 (36)
	Bowman	BW419	8.571 (73)	7.008 (49)
		BW902	6.988 (49)	5.967 (36)
Yes	Barke	BW419	6.939 (48)	5.792 (34)
		BW902	6.862 (47)	6.459 (42)
	Bowman	BW419	7.944 (63)	7.042 (50)
		BW902	7.1 (50)	5.991 (36)

Appendix 10. Factors significantly affecting mean spike length across the Bowman and Barke F2 populations

		Mean spike length (cm)
Parent(p<0.001)	Barke	9.2
	Bowman	8.7
Mutant allele (p<0.001)	BW419	9.1
	BW902	8.8
Row-type (p<0.001)	Two-rowed	9.1
	vrs3 mutant	8.8
Resown (p=0.009)	No	8.9
	Yes	9.1

Appendix 11. SED values for the pairwise transformed mean combinations. Significant differences between genotypic pair means are highlighted in grey.

:

Cross	VRS 1	INT-C		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
BW419Mo	2	2	1	*																
BW419Mo	2	6	2	0.20	*															
BW419Mo	2	H	3	0.17	0.17	*														
BW419Mo	6	2	4	0.20	0.20	0.17	*													
BW419Mo	6	6	5	0.20	0.20	0.17	0.20	*												
BW419Mo	6	H	6	0.17	0.17	0.14	0.18	0.17	*											
BW419Mo	H	2	7	0.18	0.17	0.14	0.18	0.17	0.15	*										
BW419Mo	H	6	8	0.18	0.18	0.15	0.19	0.18	0.15	0.16	*									
BW419Mo	H	H	9	0.15	0.15	0.11	0.16	0.15	0.12	0.12	0.13	*								
BW902Mo	2	2	10	0.22	0.21	0.19	0.22	0.21	0.19	0.20	0.20	0.17	*							
BW902Mo	2	6	11	0.20	0.20	0.17	0.20	0.20	0.17	0.17	0.18	0.15	0.21	*						
BW902Mo	2	H	12	0.17	0.17	0.13	0.17	0.17	0.13	0.14	0.15	0.11	0.19	0.17	*					
BW902Mo	6	2	13	0.18	0.18	0.14	0.18	0.18	0.15	0.15	0.16	0.12	0.20	0.18	0.14	*				
BW902Mo	6	6	14	0.21	0.21	0.18	0.21	0.21	0.18	0.19	0.19	0.16	0.22	0.21	0.18	0.19	*			
BW902Mo	6	H	15	0.17	0.17	0.13	0.17	0.17	0.14	0.14	0.15	0.11	0.19	0.17	0.13	0.14	0.18	*		
BW902Mo	H	2	16	0.17	0.17	0.14	0.18	0.17	0.14	0.14	0.15	0.11	0.19	0.17	0.13	0.15	0.18	0.14	*	
BW902Mo	H	6	17	0.19	0.18	0.15	0.19	0.18	0.15	0.16	0.17	0.13	0.20	0.18	0.15	0.16	0.19	0.15	0.15	*
BW902Mo	H	H	18	0.16	0.15	0.11	0.16	0.15	0.12	0.13	0.13	0.09	0.18	0.15	0.11	0.13	0.17	0.11	0.12	0.14

Appendix 12. **Vrs3** ***primer sequences***

Primer name	Primer sequence (5'-3')	Position (bp from predicted ATG start codon)
c2 1F	TTTCGGTTCCACCTATGACC	1136
c2 1R	AATACACTGGCAAGAAAAATGG	1702
c2 2F	AAACTGTGGGGTCAGATTGC	1448
c2 2R	ATTTCACGCCAGAATTCCTC	2097
c2 3F	TGGACGGTGCGCTACTCT	1920
c2 3R	TGATGCTGGTTATGTGATGAA	2599
c2 4F	TTCTTAATTTGGATTTGTCTTGATTG	2399
c2 4R	CTCCACAATTGAAGCCTGTG	2973
c2 5F	GTTTGCATGGCATGTTGAAG	2491
c2 5R	TGGATGTTCAGAATTGATTGTTTT	3142
c2 6F	TTCTGCTTGGTAATTTCTGCAT	2914
c2 6R	TTCAGGGGAAGGCTAGAGAA	3394
c2 7F	TGATCATGGTTATTCCTCATTCC	2550
c2 7R	TTCGCATCAACCTCACAAAG	3192
c2 8F	AAAACAATCAATTCTGAACATCCA	3142
c2 8R	AAGATGTCAACGATGAAGTTGC	3719
c2 9F	TGAAGCAAACTTGCTCGAAG	3667
c2 9R	TCCATGGATTCTCAGTTGGTT	4346
c2 10F	GTGAACTGCTGCTTCCTTGC	4248
c2 10R	TCGCCTTCGTGTAGTTAATGC	4927
c2 11F	ATCTACCGGCAGCTGTTCC	4669
c2 11R	AGACACCATCCCGTCTCAAC	5344
c2 12F	CTCGTAGCGCTGCAAACAG	4761
c2 12R	CCTCATCTGTAACGCATGAAT	5440
c2 14F	ATCACCCTCCCCACCTCCCCC	-354
c2 14R	CAATTGGCGTGGTTCGCGGC	224
c2 15F	TTTCTCCGTCCACTTCCGCCG	71
c2 15R	ATGCAACCCCACACGGCAAGG	630
c2 16F	CACCAGCTGCGCTTCTACCCC	505
c2 16R	GCTTCCCTGCCACGACCATCC	1036
c2 17F	GCAGCGAGGGTTTCTACGAATTGC	919
c2 17R	CCCCACAGTTTGCTGGAGTCCG	1437

Appendix 13. Vrs3 sequencing PCR conditions

65°C-60°C touchdown -primers 1, 2, 3, 7, 8, 9, 11

95°C: 15minutes
7 cycles:
94°C: 30 seconds
65°C (-0.8°C per cycle): 30 seconds
72°C: 30 seconds
30 cycles:
94°C: 30 seconds
60°C: 30 seconds
72°C: 30 seconds
72°C: 7 minutes

63°C-54°C touchdown -primers 15, 16, 17

95°C: 15minutes
10 cycles:
94°C: 30 seconds
63°C (-1.0°C per cycle): 30 seconds
72°C: 30 seconds
30 cycles:
94°C: 30 seconds
54°C: 30 seconds
72°C: 30 seconds
72°C: 7 minutes

65°C-57°C Touchdown -primer 14

95°C: 15minutes
10 cycles:
94°C: 30 seconds
65°C (-0.8°C per cycle): 30 seconds
72°C: 30 seconds
30 cycles:
94°C: 30 seconds
57°C: 30 seconds
72°C: 30 seconds
72°C: 7 minutes

62°C-57°C Touchdown -primer 5 and 12

95°C: 15minutes
7 cycles:
94°C: 30 seconds
62°C (-0.8°C per cycle): 30 seconds
72°C: 30 seconds
30 cycles:
94°C: 30 seconds
57°C: 30 seconds

72°C: 30 seconds

72°C: 7 minutes

52°C Standard -primers 4&6

95°C: 15minutes
30 cycles:
94°C: 30 seconds
52°C: 60 seconds
72°C: 30 seconds
72°C: 5 minutes

62°C Standard-primer 10

95°C: 15minutes
30 cycles:
94°C: 30 seconds
62°C: 60 seconds
72°C: 30 seconds
72°C: 5 minutes

Appendix 14.* Vrs3.*w*/Vrs3.*x* *KASP assay design nucleotide sequence

GGTCCATCAGTTGGCACTGACCAGGACATAAGCAACAGTGAAGCAAACTTGCTCGAAGC
AAATGCTGCTGACTGTGGAAAGAGTTCTCCTGCAACTTCATCGTTGACATCTTTTGCAC
TTCGTGACGGGTCTCTGCCTGCAGAACCAAAGGTCAGTTTACCACACTTGTATGAAATC
TTGTTTCCAACTGTT[A/G]TCTTTCAGTTTGAGATAAACGATATTGTTGTTTTTCAGG
TCCATGCAGCTCGAACCGACCAAATTTGGTCAATTGCTAAGCA

Appendix 15. BW419 and BW902 diagnostic KASP design

BW419

TCTGCCTTCATCATGGTAGGAAATCCTGACATTGCTGTTTCTCTAGCCTTCCCCTGAAGTTTGT
AATTTATCTGATTGTTTGTTGATTGGAGTACAGAACAAGAACT[GC/--]GAGCTGCTCTTGTA
AATCTGACCGGATTGTCTACGTCAGAGAAGACATACTGGAGTTAGAGGCTATATATAGAAAATT
CGAGCAGGATATTCGCCTGGATAAGGAAACAAGTGCTAATATCTCGTATAAGCAAGCTGCGATT
TCTGATATTGGTGTCGATCATGGTCCATCAGTTGGCA

BW902

CGGTACACCTTCCTTTTACCTCAAACCCATTTTTCTTGCCAGTGTATTACCGAATGGATGAATA
AATAATGACCATGTCATCTTTGTCTCCAGGCATT[-/T]GCAAAATTGTGGCTCCTGTAAGCG
CTTCTGTGCCTGCTGGTGTCGTGCTGATGAAGGAACAACCTGGTTTTAAGTTCATGACTAGAGT
TCAGCCGCTTCGTCTTGCTGAATGGGCTGAAGATGATACGGTCAC

Appendix 16. VRS3 splice site mutants predicted protein sequence

VRS3 Functional Domains are highlighted turquoise (JmjN), dark blue (JmjC) and green (ZnF), non synonymous amino acids are highlighted in red. The grey italicised amino acids illustrate the four amino acids deleted in the use of the int-a.59 alternative 5' splice site.

VRS3 Wild-type

MVEGRNCLPAEVRNGLETLKKRRLERMRLSAQKDEGDNPAMAARSGGDSLRTPANCGVRLHANNVTEGLP
STSSAQNNNPFAKRKVDKFNMSNLEWIDSIPECPVYCPTKEEFEDPVAYIQKISPVASKYGICKIVAPVS
ASVPAGVVLMKEQPGFKFMTRVQPLRLAEWAEDDTVTFFMSGRKYTFRDYERMANKVFSKKYSSASCLPA
RYVEEEFWREISSGKMDFVEYACDVDGSAFSSSSRDQLGKSNWNLKNFSRLPSSVLRLLQTPIPGVTDPM
LYIGMLFSMFAWHVEDHYLYSINYHHCGAFKTWYGIPGDAAPGFEKVASQYVYNKDILTGDGEDAAFDVL
LGKTTMFPPNILLDHNVPVYKAVQKPGEFVITFPRSYHSGFSHGFNCGEAVNFAIGDWFPLGSLASKRYA
LLNRTPFLAHEELLCLSAMLLSHKLSDPKTINSEHPYTQYCVKSSFVRLMRMQRRTRSLLAKMGSQIYYK
PKMYSNLSCSMCRRDCYVTHVSCGCTFDPICLHHEQELRSCSCKSDRIVYVREDILELEAIYRKFEQDIR
LDKETSANISYKQAAISDIGVDHGPSVGTDQDISNSEANLLEANAADCGKSSPATSSLTSFALRDGSLPA
EPKVHAARTDQIWSIAKQAIKTSSVEGNGALDGNSSCMADACNEISSCNASPMEYSGNSDSDSEIFRVKR
RSSILGRSAPDTKTTNLSEQKVLKRLKEASPETQHENKRHEEDSERTSVPSVSRRHNKSNSGSSEEDRED
MVPIAWRMKRRQLEAQQGDTSYAALQSKAYPSTGSCSRQQFAEATKDAASEVRPKRVKIRLPRSAANRLV
EGRQQQVSSGQGFAVDDKPPGFWHTV*

INT-A.88 Intron 3 retention

MVEGRNCLPAEVRNGLETLKKRRLERMRLSAQKDEGDNPAMAARSGGDSLRTPANCGVRLHANNVTEGLP
STSSAQNNNPFAKRKVDKFNMSNLEWIDSIPECPVYCPTKEEFEDPVAYIQKISPVASKYGICKIVAPVS
ASVPAGVVLMKEQPGFKFMTRVQPLRLAEWAEDDTVTFFMSGRCATLCLFQYLFS*

INT-A.88 Alternative 5' splice site

MVEGRNCLPAEVRNGLETLKKRRLERMRLSAQKDEGDNPAMAARSGGDSLRTPANCGVRLHANNVTEGLP
STSSAQNNNPFAKRKVDKFNMSNLEWIDSIPECPVYCPTKEEFEDPVAYIQKISPVASKYGICKIVAPVS
ASVPAGVVLMKEQPGFKFMTRVQPLRLAEWAEDDTKVHFPRLRENGQQSVL*

INT-A.59 Intron 7 retention

MVEGRNCLPAEVRNGLETLKKRRLERMRLSAQKDEGDNPAMAARSGGDSLRTPANCGVRLHANNVTEGLP
STSSAQNNNPFAKRKVDKFNMSNLEWIDSIPECPVYCPTKEEFEDPVAYIQKISPVASKYGICKIVAPVS
ASVPAGVVLMKEQPGFKFMTRVQPLRLAEWAEDDTVTFFMSGRKYTFRDYERMANKVFSKKYSSASCLPA
RYVEEEFWREISSGKMDFVEYACDVDGSAFSSSSRDQLGKSNWNLKNFSRLPSSVLRLLQTPIPGVTDPM
LYIGMLFSMFAWHVEDHYLYSINYHHCGAFKTWYGIPGDAAPGFEKVASQYVYNKDILTGDGEDAAFDVL
LGKTTMFPPNILLDHNVPVYKAVQKPGEFVITFPRSYHSGFSHDMYSFFYLLLGNFCIYAYYIF*

INT-A.59 alternative 5' splice site

MVEGRNCLPAEVRNGLETLKKRRLERMRLSAQKDEGDNPAMAARSGGDSLRTPANCGVRLHANNVTEGLP
STSSAQNNNPFAKRKVDKFNMSNLEWIDSIPECPVYCPTKEEFEDPVAYIQKISPVASKYGICKIVAPVS
ASVPAGVVLMKEQPGFKFMTRVQPLRLAEWAEDDTVTFFMSGRKYTFRDYERMANKVFSKKYSSASCLPA
RYVEEEFWREISSGKMDFVEYACDVDGSAFSSSSRDQLGKSNWNLKNFSRLPSSVLRLLQTPIPGVTDPM
LYIGMLFSMFAWHVEDHYLYSINYHHCGAFKTWYGIPGDAAPGFEKVASQYVYNKDILTGDGEDAAFDVL
LGKTTMFPPNILLDHNVPVYKAVQKPGEFVITFPRSYHSGFSHGFNCGEAVNFAIGDWFPLGSLASKRYA
LLNRTPFLAHEELLCLSAMLLSHKLSDPKTINSEHPYTQYCVKSSFVRLMRMQRRTRSLLAKMGSQIYYK
PKMYSNLSCSMCRRDCYVTHVSCGCTFDPICLHHEQELRSCSCKSDRIVYVREDILELEAIYRKFEQDIR
LDKETSANISYKQAAISDIGVDHGPSVGTDQDISNSEANLLEANAADCGKSSPATSSLTSFALRDGSLPA
EPKVHAARTDQIWSIAKQAIKTSSVEGNGALDGNSSCMADACNEISSCNASPMEYSGNSDSDSEIFRVKR
RSSILGRSAPDTKTTNLSEQKVLKRLKEASPETQHENKRHEEDSERTSVPSVSRRHNKSNSGSSEEDRED
MVPIAWRMKRRQLEAQQGDTSYAALQSKAYPSTGSCSRQQFAEATKDAASEVRPKRVKIRLPRSAANRLV
EGRQQQVSSGQGFAVDDKPPGFWHTV*

Appendix 17. Germplasm collection used for Vrs3 genotyping and GWAS

Year	*Vrs3* Haplotype	Row Type	Habit	Cultivar	Origin	*Vrs3* Allele	GWAS Panel

1987	1	2	SB	Alexis	Germany	*Vrs3.w*	Y
1998	1	2	SB	Alliot	Denmark	*Vrs3.w*	Y
1979	1	2	SB	Alva	Sweden	*Vrs3.w*	Y
1972	1	2	SB	Ametyst	Czech	*Vrs3.w*	Y
1980	1	2	SB	Anni	Estonia	*Vrs3.w*	Y
1996	1	2	SB	Ansis	Latvia	*Vrs3.w*	Y
1984	1	2	SB	Apex	Netherlands	*Vrs3.w*	Y
1966	1	2	SB	Arvo	Finland	*Vrs3.w*	Y
1975	1	2	SB	Athos	France	*Vrs3.w*	Y
1996	1	2	SB	Atribut	Czech	*Vrs3.w*	Y
2002	1	2	SB	Auriga	Germany	*Vrs3.w*	Y
1942	1	2	SB	Balder	Sweden	*Vrs3.w*	Y
1990	1	2	SB	Balga	Latvia	*Vrs3.w*	Y
2005	1	2	SB	Barabas	Denmark	*Vrs3.w*	Y
1997	1	2	SB	Barke	Germany	*Vrs3.w*	Y
1989	1	2	SB	Baronesse	Germany	*Vrs3.w*	Y
1972	1	2	SB	Berenice	France	*Vrs3.w*	Y
1916	1	2	SB	Binder	Denmark	*Vrs3.w*	Y
1963	1	2	SB	Birgitta	Sweden	*Vrs3.w*	Y
1981	1	2	SB	Birka	Sweden	*Vrs3.w*	Y
1984	1	2	SB	Bowman	North America	*Vrs3.w*	N
1999	1	2	SB	Braemar	UK	*Vrs3.w*	Y
1964	1	2	SB	Britta	Sweden	*Vrs3.w*	Y
2004	1	2	SB	Cabaret	Denmark	*Vrs3.w*	Y
1980	1	2	SB	Caja	Denmark	*Vrs3.w*	Y
2002	1	2	SB	Calgary	France	*Vrs3.w*	Y
1985	1	2	SB	Cameo	UK	*Vrs3.w*	Y
1994	1	2	SB	Caminant	Denmark	*Vrs3.w*	Y
1985	1	2	SB	Camir	Denmark	*Vrs3.w*	Y
1998	1	2	SB	Cellar	UK	*Vrs3.w*	Y
1987	1	2	SB	Cheri	Germany	*Vrs3.w*	Y
1999	1	2	SB	Cicero	Denmark	*Vrs3.w*	Y
1980	1	2	SB	Claret	UK	*Vrs3.w*	Y
2006	1	2	SB	Claude	Denmark	*Vrs3.w*	Y
1991	1	2	SB	Cooper	UK	*Vrs3.w*	Y
1982	1	2	SB	Croydon	Sweden	*Vrs3.w*	Y
1984	1	2	SB	Dandy	UK	*Vrs3.w*	Y
2000	1	2	SB	Danuta	Germany	*Vrs3.w*	Y
1965	1	2	SB	Deba Abed	Denmark	*Vrs3.w*	Y
1959	1	2	SB	Delta	Netherlands	*Vrs3.w*	Y
2000	1	2	SB	Dialog	Denmark	*Vrs3.w*	Y
1965	1	2	SB	Diamant	Czech	*Vrs3.w*	Y
1991	1	2	SB	Digersano	Italy	*Vrs3.w*	Y
1975	1	2	SB	Dina	Denmark	*Vrs3.w*	Y
1985	1	2	SB	Doublet	UK	*Vrs3.w*	Y
1968	1	2	SB	Drake	Netherlands	*Vrs3.w*	Y
1954	1	2	SB	Drost	Denmark	*Vrs3.w*	Y

1980	1	2	SB	Egmont	UK	*Vrs3.w*	Y
1998	1	2	SB	Elantra	Denmark	*Vrs3.w*	Y
1989	1	2	SB	Elo	Estonia	*Vrs3.w*	Y
1962	1	2	SB	Emir	Netherlands	*Vrs3.w*	Y
2000	1	2	SB	Eunova	Austria	*Vrs3.w*	Y
1996	1	2	SB	Famin	Czech	*Vrs3.w*	Y
1973	1	2	SB	Favorit	Czech	*Vrs3.w*	Y
2002	1	2	SB	Felicitas	Germany	*Vrs3.w*	Y
1987	1	2	SB	Formula	Sweden	*Vrs3.w*	Y
1993	1	2	SB	Forum	Czech	*Vrs3.w*	Y
1941	1	2	SB	Freja	Sweden	*Vrs3.w*	Y
1990	1	2	SB	Galan	Slovakia	*Vrs3.w*	Y
1997	1	2	SB	Gant	Denmark	*Vrs3.w*	Y
1995	1	2	SB	Gate	Latvia	*Vrs3.w*	Y
1970	1	2	SB	Gerkra	Netherlands	*Vrs3.w*	Y
1976	1	2	SB	Gitane	Netherlands	*Vrs3.w*	Y
2006	1	2	SB	Gizmo	Denmark	*Vrs3.w*	Y
1980	1	2	SB	Golf	UK	*Vrs3.w*	Y
1973	1	2	SB	Hana	Czech	*Vrs3.w*	Y
X	1	2	SB	Hanka	Germany	*Vrs3.w*	Y
1977	1	2	SB	Hanna	Sweden	*Vrs3.w*	Y
1971	1	2	SB	Hassan	Netherlands	*Vrs3.w*	Y
1967	1	2	SB	Hellas	Sweden	*Vrs3.w*	Y
1942	1	2	SB	Helmi	Finland	*Vrs3.w*	Y
1998	1	2	SB	Heris	Czech	*Vrs3.w*	Y
1999	1	2	SB	Hydrogen	Denmark	*Vrs3.w*	Y
1980	1	2	SB	Ida	Sweden	*Vrs3.w*	Y
1970	1	2	SB	Imber	UK	*Vrs3.w*	Y
X	1	2	SB	Impala	Netherlands	*Vrs3.w*	Y
1956	1	2	SB	Ingrid	Sweden	*Vrs3.w*	Y
1924	1	2	SB	Isaria	Germany	*Vrs3.w*	Y
1987	1	2	SB	Jarek	Czech	*Vrs3.w*	Y
1981	1	2	SB	Karat	Czech	*Vrs3.w*	Y
1931	1	2	SB	Kenia	Denmark	*Vrs3.w*	Y
2006	1	2	SB	Keops	Denmark	*Vrs3.w*	Y
2002	1	2	SB	Kristaps	Latvia	*Vrs3.w*	Y
1991	1	2	SB	Krona	Germany	*Vrs3.w*	Y
1981	1	2	SB	Krystal	Czech	*Vrs3.w*	Y
1976	1	2	SB	Lud	UK	*Vrs3.w*	Y
1997	1	2	SB	Lux	Denmark	*Vrs3.w*	Y
1994	1	2	SB	Lysimax	Denmark	*Vrs3.w*	Y
2001	1	2	SB	Malva	Latvia	*Vrs3.w*	Y
1986	1	2	SB	Maresi	Germany	*Vrs3.w*	Y
1973	1	2	SB	Maris Mink	UK	*Vrs3.w*	Y
1983	1	2	SB	Mars	Czech	*Vrs3.w*	Y
2005	1	2	SB	Marthe	Germany	*Vrs3.w*	Y
2004	1	2	SB	Mauritia	Germany	*Vrs3.w*	Y

1993	1	2	SB	Mentor	Sweden	*Vrs3.w*	Y
1970	1	2	SB	Midas	UK	*Vrs3.w*	Y
2006	1	2	SB	Nathalie	Denmark	*Vrs3.w*	Y
2002	1	2	SB	NFC Tipple	UK	*Vrs3.w*	Y
1971	1	2	SB	Nordal	Sweden	*Vrs3.w*	Y
1988	1	2	SB	Novum	Slovakia	*Vrs3.w*	Y
2003	1	2	SB	Odessa	Germany	*Vrs3.w*	Y
1975	1	2	SB	Okos	Italy	*Vrs3.w*	Y
1986	1	2	SB	Orbit	Slovakia	*Vrs3.w*	Y
1972	1	2	SB	Otto	Italy	*Vrs3.w*	Y
1996	1	2	SB	Paloma	Denmark	*Vrs3.w*	Y
1998	1	2	SB	Pasadena	Germany	*Vrs3.w*	Y
1988	1	2	SB	Perun	Czech	*Vrs3.w*	Y
1995	1	2	SB	Primus	Czech	*Vrs3.w*	Y
1953	1	2	SB	Proctor	UK	*Vrs3.w*	Y
1996	1	2	SB	Prosa	Austria	*Vrs3.w*	Y
2004	1	2	SB	Publican	UK	*Vrs3.w*	Y
1992	1	2	SB	Quartz	UK	*Vrs3.w*	Y
2004	1	2	SB	Quench	UK	*Vrs3.w*	Y
1976	1	2	SB	Rapid	Czech	*Vrs3.w*	Y
1991	1	2	SB	Rasa	Latvia	*Vrs3.w*	Y
1951	1	2	SB	Rika	Sweden	*Vrs3.w*	Y
1992	1	2	SB	Riviera	UK	*Vrs3.w*	Y
1982	1	2	SB	Roland	Sweden	*Vrs3.w*	Y
1999	1	2	SB	Roxana	Germany	*Vrs3.w*	Y
1982	1	2	SB	Rubin	Czech	*Vrs3.w*	Y
1978	1	2	SB	Safir	Czech	*Vrs3.w*	Y
1973	1	2	SB	Salka	Denmark	*Vrs3.w*	Y
1974	1	2	SB	Senat	Sweden	*Vrs3.w*	Y
1995	1	2	SB	Sencis	Latvia	*Vrs3.w*	Y
2003	1	2	SB	Simba	Denmark	*Vrs3.w*	Y
1978	1	2	SB	Simon	Sweden	*Vrs3.w*	Y
2004	1	2	SB	Smilla	Denmark	*Vrs3.w*	Y
1977	1	2	SB	Spartan	Czech	*Vrs3.w*	Y
1996	1	2	SB	Static	UK	*Vrs3.w*	Y
1989	1	2	SB	Steffi	Germany	*Vrs3.w*	Y
1968	1	2	SB	Sultan	Netherlands	*Vrs3.w*	Y
2005	1	2	WB	AC99/077/2	Germany	*Vrs3.w*	Y
1994	1	2	WB	Aci	Italy	*Vrs3.w*	Y
1993	1	2	WB	Alfeo	Italy	*Vrs3.w*	Y
1972	1	2	WB	Alpha	France	*Vrs3.w*	Y
1995	1	2	WB	Amillis	France	*Vrs3.w*	Y
1985	1	2	WB	Arda	Italy	*Vrs3.w*	Y
1993	1	2	WB	Asso	Italy	*Vrs3.w*	Y
2004	1	2	WB	Blythe	Sweden	*Vrs3.w*	Y
2000	1	2	WB	Boreale	France	*Vrs3.w*	Y
2002	1	2	WB	Calliope	Sweden	*Vrs3.w*	Y

1999	1	2	WB	Carat	UK	*Vrs3.w*	Y
2002	1	2	WB	Concept	UK	*Vrs3.w*	Y
2001	1	2	WB	Connoisseur	UK	*Vrs3.w*	Y
2002	1	2	WB	Coriolis	UK	*Vrs3.w*	Y
2002	1	2	WB	CPBT_B66	UK	*Vrs3.w*	Y
2005	1	2	WB	CPBT_B78	UK	*Vrs3.w*	Y
1999	1	2	WB	Diadem	UK	*Vrs3.w*	Y
1999	1	2	WB	Diamond	UK	*Vrs3.w*	Y
1993	1	2	WB	Druid	UK	*Vrs3.w*	Y
1992	1	2	WB	Electron	UK	*Vrs3.w*	Y
1992	1	2	WB	Fanfare	UK	*Vrs3.w*	Y
1988	1	2	WB	Finesse	UK	*Vrs3.w*	Y
1992	1	2	WB	Gazelle	UK	*Vrs3.w*	Y
1993	1	2	WB	Gleam	UK	*Vrs3.w*	Y
1994	1	2	WB	Glint	UK	*Vrs3.w*	Y
1996	1	2	WB	Goldrush	Sweden	*Vrs3.w*	Y
1987	1	2	WB	Gypsy	UK	*Vrs3.w*	Y
1982	1	2	WB	Halcyon	UK	*Vrs3.w*	Y
2002	1	2	WB	Hermia	UK	*Vrs3.w*	Y
1995	1	2	WB	Honey	UK	*Vrs3.w*	Y
1997	1	2	WB	Hurricane	UK	*Vrs3.w*	Y
2002	1	2	WB	Imogen	Sweden	*Vrs3.w*	Y
1981	1	2	WB	Kaskade	Germany	*Vrs3.w*	N
1995	1	2	WB	Lark	UK	*Vrs3.w*	Y
1997	1	2	WB	Leonie	Germany	*Vrs3.w*	Y
1997	1	2	WB	Mariner	UK	*Vrs3.w*	Y
1997	1	2	WB	Masai	UK	*Vrs3.w*	Y
2000	1	2	WB	Mead	UK	*Vrs3.w*	Y
2002	1	2	WB	MHHX011	France	*Vrs3.w*	Y
1993	1	2	WB	Mystique	UK	*Vrs3.w*	Y
2000	1	2	WB	Parasol	UK	*Vrs3.w*	Y
1995	1	2	WB	Peridot	UK	*Vrs3.w*	Y
1999	1	2	WB	Pippa	Denmark	*Vrs3.w*	Y
1998	1	2	WB	Platine	France	*Vrs3.w*	Y
1994	1	2	WB	Portrait	UK	*Vrs3.w*	Y
1993	1	2	WB	Prelude	UK	*Vrs3.w*	Y
2002	1	2	WB	Rattle	UK	*Vrs3.w*	Y
1995	1	2	WB	Ravel	UK	*Vrs3.w*	Y
1992	1	2	WB	Rhythm	UK	*Vrs3.w*	Y
1994	1	2	WB	Rifle	UK	*Vrs3.w*	Y
2000	1	2	WB	Saigon	UK	*Vrs3.w*	Y
1998	1	2	WB	Sonic	UK	*Vrs3.w*	Y
1995	1	2	WB	Spirit	UK	*Vrs3.w*	Y
1988	1	2	WB	Sprite	UK	*Vrs3.w*	Y
1998	1	2	WB	Sumo	UK	*Vrs3.w*	Y
1992	1	2	WB	Sunrise	UK	*Vrs3.w*	Y
2004	1	2	WB	SW_165	Sweden	*Vrs3.w*	Y

2003	1	2	WB	SW Norma	Sweden	*Vrs3.w*	Y
2000	1	2	WB	Swallow	UK	*Vrs3.w*	Y
1990	1	2	WB	Swift	UK	*Vrs3.w*	Y
1992	1	2	WB	Tempo	UK	*Vrs3.w*	Y
2002	1	2	WB	Thalia	UK	*Vrs3.w*	Y
1999	1	2	WB	Tipster	UK	*Vrs3.w*	Y
1995	1	2	WB	Toffee	UK	*Vrs3.w*	Y
1993	1	2	WB	Tokyo	UK	*Vrs3.w*	Y
1992	1	2	WB	Tosca	UK	*Vrs3.w*	Y
1995	1	2	WB	Turine	France	*Vrs3.w*	Y
1993	1	2	WB	Vanilla	France	*Vrs3.w*	Y
1994	1	2	WB	Volley	UK	*Vrs3.w*	Y
1998	1	2	WB	Weaver	Sweden	*Vrs3.w*	Y
1997	1	2	WB	Wizard	UK	*Vrs3.w*	Y
1997	1	2	WB	Zulu	UK	*Vrs3.w*	Y
1978	2	6	SB	Morex	USA	*Vrs3.x*	Y
2003	2	2	WB	AC97/H2406/10	Germany	*Vrs3.x*	Y
2000	2	2	WB	Antelope	UK	*Vrs3.x*	Y
1994	2	2	WB	Antigua	UK	*Vrs3.x*	Y
1997	2	2	WB	Antonia	France	*Vrs3.x*	Y
2001	2	2	WB	Aquarelle	Germany	*Vrs3.x*	Y
2003	2	2	WB	Archimedes	UK	*Vrs3.x*	Y
1997	2	2	WB	Artist	UK	*Vrs3.x*	Y
1998	2	2	WB	Avenue	UK	*Vrs3.x*	Y
1998	2	2	WB	Barcelona	Netherlands	*Vrs3.x*	Y
1995	2	2	WB	Baton	UK	*Vrs3.x*	Y
1996	2	2	WB	Bistro	UK	*Vrs3.x*	Y
1996	2	2	WB	Breeze	UK	*Vrs3.x*	Y
1989	2	2	WB	Bronze	UK	*Vrs3.x*	Y
2000	2	2	WB	Calcutta	UK	*Vrs3.x*	Y
1998	2	2	WB	Campion	UK	*Vrs3.x*	Y
1995	2	2	WB	Candy	UK	*Vrs3.x*	Y
2004	2	2	WB	CEBECO 02215/05	Netherlands	*Vrs3.x*	N
2004	2	2	WB	Cedar	UK	*Vrs3.x*	Y
2000	2	2	WB	Cellina	Netherlands	*Vrs3.x*	Y
2003	2	2	WB	Celsius	UK	*Vrs3.x*	Y
1998	2	2	WB	Chamomile	UK	*Vrs3.x*	Y
1990	2	2	WB	Chestnut	UK	*Vrs3.x*	Y
1999	2	2	WB	Chicane	UK	*Vrs3.x*	Y
1996	2	2	WB	Chintz	UK	*Vrs3.x*	Y
1996	2	2	WB	Chord	UK	*Vrs3.x*	Y
2003	2	2	WB	Cinnamon	UK	*Vrs3.x*	Y
2000	2	2	WB	Clara_2RWB2	Germany	*Vrs3.x*	Y
1994	2	2	WB	Cobalt	UK	*Vrs3.x*	Y
1994	2	2	WB	Credo	Netherlands	*Vrs3.x*	Y

1995	2	2	WB	Crescendo	UK	*Vrs3.x*	Y
1998	2	2	WB	Cynthia	UK	*Vrs3.x*	Y
2003	2	2	WB	Cypress	UK	*Vrs3.x*	Y
2003	2	2	WB	Dolphin	UK	*Vrs3.x*	Y
1992	2	2	WB	Duet	UK	*Vrs3.x*	Y
1992	2	2	WB	Emeraude	France	*Vrs3.x*	Y
2003	2	2	WB	Fahrenheit	UK	*Vrs3.x*	Y
1989	2	2	WB	Frolic	UK	*Vrs3.x*	Y
1998	2	2	WB	Goldmine	UK	*Vrs3.x*	Y
1997	2	2	WB	Halifax	UK	*Vrs3.x*	Y
1990	2	2	WB	Intro	Netherlands	*Vrs3.x*	Y
1994	2	2	WB	Jet	UK	*Vrs3.x*	Y
1989	2	2	WB	Karisma	UK	*Vrs3.x*	Y
1990	2	2	WB	Kelibia	France	*Vrs3.x*	Y
2001	2	2	WB	Kingston	UK	*Vrs3.x*	Y
1985	2	2	WB	Kira	Germany	*Vrs3.x*	Y
1997	2	2	WB	Laurel	UK	*Vrs3.x*	Y
2001	2	2	WB	Louise	Denmark	*Vrs3.x*	Y
1992	2	2	WB	Madrigal	UK	*Vrs3.x*	Y
1996	2	2	WB	Mahogany	UK	*Vrs3.x*	Y
2001	2	2	WB	Maritem	Germany	*Vrs3.x*	Y
1993	2	2	WB	Medoc	UK	*Vrs3.x*	Y
2002	2	2	WB	Merode	Netherlands	*Vrs3.x*	Y
1999	2	2	WB	Milena	Germany	*Vrs3.x*	Y
1998	2	2	WB	Montage	UK	*Vrs3.x*	Y
1999	2	2	WB	Moonshine	UK	*Vrs3.x*	Y
1996	2	2	WB	Musette	Netherlands	*Vrs3.x*	Y
2000	2	2	WB	Outlook	UK	*Vrs3.x*	Y
1980	2	2	WB	Panda_2RWB	France	*Vrs3.x*	N
2000	2	2	WB	Pedigree	UK	*Vrs3.x*	Y
1994	2	2	WB	Pilot	UK	*Vrs3.x*	Y
1986	2	2	WB	Posaune	Germany	*Vrs3.x*	Y
1993	2	2	WB	Punch	UK	*Vrs3.x*	Y
2004	2	2	WB	Retriever	Denmark	*Vrs3.x*	Y
1999	2	2	WB	Scylla	UK	*Vrs3.x*	Y
1996	2	2	WB	Sevilla	Netherlands	*Vrs3.x*	Y
2000	2	2	WB	Sombrero	UK	*Vrs3.x*	Y
1999	2	2	WB	Spinner	Sweden	*Vrs3.x*	Y
2001	2	2	WB	SW_Alison	Sweden	*Vrs3.x*	Y
2000	2	2	WB	SW_Farrier	Sweden	*Vrs3.x*	Y
1996	2	2	WB	Tabetha	UK	*Vrs3.x*	Y
1999	2	2	WB	Tucker	UK	*Vrs3.x*	Y
1993	2	2	WB	Tudor	UK	*Vrs3.x*	Y
2000	2	2	WB	Verticale	France	*Vrs3.x*	Y
1997	2	2	WB	Vilna	Netherlands	*Vrs3.x*	Y
2001	2	2	WB	Wombat	UK	*Vrs3.x*	Y
1977	2	6	WB	Athene	Germany	*Vrs3.x*	Y

1996	2	6	WB	Balaki	France	*Vrs3.x*	Y
1998	2	6	WB	Carola	Germany	*Vrs3.x*	Y
1953	2	6	WB	Dea	Germany	*Vrs3.x*	Y
X	2	6	WB	Esterel	France	*Vrs3.x*	Y
1994	2	6	WB	Federal	France	*Vrs3.x*	Y
1980	2	6	WB	Franka	Germany	*Vrs3.x*	Y
2006	2	6	WB	Fridericus	Germany	*Vrs3.x*	Y
1988	2	6	WB	Frost	Sweden	*Vrs3.x*	Y
1977	2	6	WB	Gerbel	France	*Vrs3.x*	Y
1988	2	6	WB	Glenan	France	*Vrs3.x*	Y
1989	2	6	WB	Grete	Germany	*Vrs3.x*	Y
X	2	6	WB	Herfordia	Germany	*Vrs3.x*	Y
2002	2	6	WB	Ketos	France	*Vrs3.x*	Y
2005	2	6	WB	Lonni	Denmark	*Vrs3.x*	Y
1997	2	6	WB	Lorena	Germany	*Vrs3.x*	Y
2004	2	6	WB	Marado	France	*Vrs3.x*	Y
1998	2	6	WB	Mattina	France	*Vrs3.x*	Y
1972	2	6	WB	Mirra	Germany	*Vrs3.x*	Y
1993	2	6	WB	Muscat	UK	*Vrs3.x*	Y
2003	2	6	WB	Naomie	Germany	*Vrs3.x*	Y
1997	2	6	WB	Passport	France	*Vrs3.x*	Y
1978	2	6	WB	Pirate	France	*Vrs3.x*	Y
1980	2	6	WB	Plaisant	France	*Vrs3.x*	Y
1989	2	6	WB	Princess	France	*Vrs3.x*	Y
1963	2	6	WB	Senta	Germany	*Vrs3.x*	Y
1992	2	6	WB	Sonora	France	*Vrs3.x*	Y
1970	3	2	SB	Akka	Sweden	*Vrs3.x*	Y
1985	3	2	SB	Alis	Denmark	*Vrs3.x*	Y
1964	3	2	SB	Balder J	Sweden	*Vrs3.x*	Y
1952	3	2	SB	Bonus	Sweden	*Vrs3.x*	Y
1947	3	2	SB	Carlsberg	Denmark	*Vrs3.x*	Y
1968	3	2	SB	Golden Promise	UK	*Vrs3.x*	Y
1915	3	2	SB	Gull	Sweden	*Vrs3.x*	Y
1967	3	2	SB	Karri	Finland	*Vrs3.x*	Y
1927	3	2	SB	Maja	Denmark	*Vrs3.x*	Y
1971	3	2	SB	Mona	Sweden	*Vrs3.x*	Y
1985	3	2	SB	Nemex	Sweden	*Vrs3.x*	Y
1958	3	2	SB	Pallas	Sweden	*Vrs3.x*	Y
1972	3	2	SB	Rupal	Sweden	*Vrs3.x*	Y
1975	3	2	SB	Salve	Sweden	*Vrs3.x*	Y
1972	3	2	SB	Stendes	Latvia	*Vrs3.x*	Y
1973	3	2	WB	Igri	Germany	*Vrs3.x*	Y
1996	3	2	WB	Magnolia	UK	*Vrs3.x*	Y
2001	3	2	WB	Nocturne	UK	*Vrs3.x*	Y
1990	3	2	WB	Willow	UK	*Vrs3.x*	Y
1994	4	2	SB	Inari	Finland	*Vrs3.w*	Y
1996	4	6	SB	Botnia	Finland	*Vrs3.w*	Y

1998	4	6	SB	Edda	Sweden	*Vrs3.w*	Y
1975	4	6	SB	Eero	Finland	*Vrs3.w*	Y
1973	4	6	SB	Hankkija 673	Finland	*Vrs3.w*	Y
1939	4	6	SB	Herse	Norway	*Vrs3.w*	Y
1977	4	6	SB	Kajsa	Finland	*Vrs3.w*	Y
1981	4	6	SB	Kilta	Finland	*Vrs3.w*	Y
1918	4	6	SB	Maskin	Norway	*Vrs3.w*	Y
1986	4	6	SB	Niina	Finland	*Vrs3.w*	Y
1959	4	6	SB	Otra	Finland	*Vrs3.w*	Y
1960	4	6	SB	Paavo	Finland	*Vrs3.w*	Y
1987	4	6	SB	Pohto	Finland	*Vrs3.w*	Y
1980	4	6	SB	Pokko	Finland	*Vrs3.w*	Y
1983	4	6	SB	Potra	Finland	*Vrs3.w*	Y
1979	4	6	SB	Silja	Finland	*Vrs3.w*	Y
X	4	6	SB	Stella	Sweden	*Vrs3.w*	Y
1999	5	2	WB	Caption	UK	*Vrs3.x*	Y
2000	5	2	WB	Cathay	UK	*Vrs3.x*	Y
2003	5	2	WB	CWB 5663/1	UK	*Vrs3.x*	N
1989	5	2	WB	Eagle	UK	*Vrs3.x*	Y
1996	5	2	WB	Flute	UK	*Vrs3.x*	Y
2000	5	2	WB	Jessica	Denmark	*Vrs3.x*	Y
2001	5	2	WB	Lambada	UK	*Vrs3.x*	Y
2001	5	2	WB	Mortimer	UK	*Vrs3.x*	Y
2002	5	2	WB	Murcie	France	*Vrs3.x*	Y
1999	5	2	WB	Sapphire	UK	*Vrs3.x*	Y
2002	5	2	WB	Spectrum	UK	*Vrs3.x*	Y
1981	5	6	WB	Mirco	Italy	*Vrs3.x*	Y
1980	5	6	WB	Onice	Italy	*Vrs3.x*	Y
1980	5	6	WB	Tapir	Netherlands	*Vrs3.x*	Y
1986	6	2	WB	Baraka	France	*Vrs3.x*	Y
1996	6	2	WB	Duchess	France	*Vrs3.x*	Y
1998	6	2	WB	Nure	Italy	*Vrs3.x*	Y
2001	6	2	WB	SW_Sienna	Sweden	*Vrs3.x*	Y
1992	6	6	WB	Canoro	Italy	*Vrs3.x*	Y
1970	6	6	WB	Hoppel	France	*Vrs3.x*	Y
1985	6	6	WB	Isa	Belgium	*Vrs3.x*	Y
2003	6	6	WB	Lutece	France	*Vrs3.x*	Y
1988	6	6	WB	Manitou	France	*Vrs3.x*	Y
1998	6	6	WB	Zoe	Germany	*Vrs3.x*	Y
2005	7	2	WB	410/3E	France	*Vrs3.w*	Y
2003	7	2	WB	Celebrity	UK	*Vrs3.w*	Y
2003	7	2	WB	Charleston	UK	*Vrs3.w*	Y
1998	7	2	WB	Harland	UK	*Vrs3.w*	Y
2003	7	2	WB	Houston	UK	*Vrs3.w*	Y
2002	7	2	WB	Nectaria	France	*Vrs3.w*	Y
1997	7	2	WB	Opal	UK	*Vrs3.w*	Y
2005	7	2	WB	Wintmalt	Germany	*Vrs3.w*	Y

1978	8	2	SB	Abava	Latvia	*Vrs3.w*	Y
1952	8	2	SB	Domen	Norway	*Vrs3.w*	Y
1981	8	2	SB	Gorm	Denmark	*Vrs3.w*	Y
1985	8	2	SB	Imula	Latvia	*Vrs3.w*	Y
1983	8	2	SB	Romi	Denmark	*Vrs3.w*	Y
1992	8	2	SB	Ruja	Latvia	*Vrs3.w*	Y
1996	8	2	SB	Saana	Finland	*Vrs3.w*	Y
2000	minor	2	SB	Brazil	France	*Vrs3.w*	Y
2006	minor	2	SB	Chanell	Denmark	*Vrs3.w*	Y
1968	minor	2	SB	Cilla	Sweden	*Vrs3.w*	Y
1964	minor	2	SB	Clara	Sweden	*Vrs3.x*	Y
1989	minor	2	SB	Haruna Nijo	Japan	*Vrs3.w*	N
1998	minor	2	SB	Lysiba	Denmark	*Vrs3.w*	Y
1990	minor	2	SB	Meltan	Sweden	*Vrs3.w*	Y
1980	minor	2	SB	Steina	Germany	*Vrs3.w*	Y
1955	minor	6	SB	Frisia	Germany	*Vrs3.x*	Y
1933	minor	6	SB	Jadar	Norway	*Vrs3.w*	Y
1960	minor	6	SB	Jarle	Norway	*Vrs3.w*	Y
1976	minor	6	SB	Rondo	Italy	*Vrs3.w*	Y
1973	minor	6	SB	Steptoe	USA	*Vrs3.x*	Y
2002	minor	2	WB	Aiace	Italy	*Vrs3.x*	Y
1998	minor	2	WB	Anvil	UK	*Vrs3.w*	Y
2001	minor	2	WB	Camion	UK	*Vrs3.x*	Y
2001	minor	2	WB	Dolmen	France	*Vrs3.w*	Y
1996	minor	2	WB	Heligan	UK	*Vrs3.x*	Y
1965	minor	2	WB	Malta	Germany	*Vrs3.w*	Y
2001	minor	2	WB	Tallica	UK	*Vrs3.w*	Y
2001	minor	2	WB	Wigwam	UK	*Vrs3.w*	Y
2001	minor	2	WB	Winner	Netherlands	*Vrs3.w*	Y
1998	minor	6	WB	Balda	Italy	*Vrs3.w*	Y
1961	minor	6	WB	Dura	Germany	*Vrs3.w*	Y
1937	minor	6	WB	Hatif de Grignon	France	*Vrs3.x*	Y
1995	minor	6	WB	Patricia	Austria	*Vrs3.w*	Y
1965	minor	6	WB	Vogelsanger Gold	Germany	*Vrs3.x*	Y

Appendix 18. Row-type GWAS peak association

Population	Marker	Chromosome	cM	-LOG10(P)	Candidate Locus
Spring	SCRI_RS_151764	1H	46.71	10.7	
Spring	12_10198	1H	51.49	16.7	*VRS3*
Spring	SCRI_RS_225431	1H	82.51	5.5	
Spring	12_31319	1H	92.35	78.9	
Spring	SCRI_RS_197337	1H	101.1	4.8	
Spring	12_10905	1H	106.2	13.6	
Spring	SCRI_RS_235860	2H	76.52	8.5	*VRS1*
Spring	SCRI_RS_229693	3H	43.91	6.2	*VRS4*
Spring	11_21358	3H	68.24	9.7	
Spring	11_21070	4H	25.85	12.2	*INT-C*
Spring	12_31313	4H	34.35	9.8	
Spring	SCRI_RS_206213	6H	55.97	4.2	
Spring	SCRI_RS_159133	6H	95.04	5.0	
Spring	SCRI_RS_162631	7H	19.74	11.8	
Spring	SCRI_RS_47197	7H	22.73	10.1	
Spring	SCRI_RS_234226	7H	61.72	4.1	
Winter	11_20724	2H	12.59	9.3	
Winter	SCRI_RS_137263	2H	77.09	13.1	*VRS1*
Winter	12_30609	3H	46.18	4.7	*VRS4*
Winter	SCRI_RS_141195	3H	103	5.2	
Winter	SCRI_RS_175038	3H	127	5.5	
Winter	12_30229	3H	151.3	4.9	
Winter	SCRI_RS_125487	4H	22.24	4.3	*INT-C*
Winter	11_10048	4H	45.68	6.4	
Winter	SCRI_RS_184107	4H	50.99	4.0	
Winter	SCRI_RS_229116	4H	113.7	4.5	
Winter	SCRI_RS_200387	7H	46.39	4.4	

Appendix 19. F_2 BW419Mo andBW902Mo glasshouse phenotype table of means

a)Central grain parameters

VRS1	*INT-C*	*VRS3*	*Number of tillers*	*Number of central grain*	*Central TGW(g)*	*Central grain area(mm^2)*	*Central grain length (mm)*	*Central grain width(mm)*
2	2	2	15.18	20.66	58.05	26.77	9.378	3.756
2	2	H	14.59	23.2	61.77	27.08	9.233	3.808
2	H	2	14.16	20.7	57.31	27.02	9.48	3.7
2	H	H	13.57	23.25	59.64	27.2	9.433	3.767
2	6	2	13.75	18.72	58.62	27.23	9.657	3.714
2	6	H	13.16	21.26	61.3	26.8	9.236	3.745
H	2	2	14.25	22.12	61.16	26.98	9.292	3.783
H	2	H	13.66	22.88	60.14	27.27	9.52	3.753
H	H	2	13.24	22.17	61.84	27.11	9.257	3.814
H	H	H	12.65	22.92	61.07	26.65	9.2	3.757
2	2	6	12.61	17.79	58.17	26.62	9.65	3.675
2	H	6	11.6	17.83	52.84	25.11	9.455	3.536
2	6	6	11.19	15.85	47.74	23.95	9.327	3.4
H	2	6	11.69	16.28	53.47	24.82	9.38	3.49
H	6	2	12.83	20.18	49.85	23.57	9.014	3.429
H	6	H	12.24	20.94	49.71	23.85	9.137	3.426
H	H	6	10.67	16.32	47.48	24.12	9.348	3.413
H	6	6	10.26	14.34	45.82	24.08	9.267	3.483
6	2	2	12.88	16.8	52.84	23.52	8.933	3.483
6	2	H	12.29	16.97	48.6	22.94	9.041	3.341
6	H	2	11.87	16.84	46.3	22.8	9.083	3.325
6	H	H	11.28	17.01	42.09	21.95	9.034	3.154
6	2	6	10.32	14.77	43.79	23.35	9.328	3.333
6	6	2	11.46	14.86	44.82	22.15	9	3.25
6	H	6	9.3	14.81	45.92	23.56	9.121	3.393
6	6	H	10.87	15.03	43.89	22.67	9.236	3.264
6	6	6	8.89	12.83	47.74	22.95	8.931	3.431
		SED	0.6552	1.037	2.976	0.6249	0.1538	0.0784
		Residual Variation	16.52	21.69	57.82	2.418	0.154	0.038

b)Lateral grain parameters

VRS1	*INT-C*	*VRS3*	*Number of lateral grain*	*Lateral TGW(g)*	*Lateral grain area(mm^2)*	*Lateral grain length(mm)*	*Lateral grain width(mm)*	*Central: lateral grain area ratio*	*Central: lateral grain length ratio*	*Central: lateral grain width ratio*
2	2	6	26.5	36.6	20.08	8.85	3	0.7551	0.9166	0.8169
2	H	6	24.32	35.68	19.73	8.777	2.955	0.7868	0.9288	0.8359
2	6	6	29.13	36.35	20.73	8.98	3.053	0.8673	0.9643	0.8996
H	2	6	30.74	39.16	20.68	8.979	3.037	0.8328	0.9572	0.8687
H	6	2	15.57	26.92	16.6	8.329	2.586	0.7055	0.9245	0.7537
H	6	H	26.79	26.08	17.02	8.595	2.542	0.7136	0.9409	0.7419
H	H	6	30.28	35.53	20.36	8.983	2.96	0.8446	0.9608	0.868
H	6	6	31.72	37.99	21.46	8.956	3.172	0.8913	0.9673	0.9107
6	2	2	27	34.33	17.52	8.1	2.85	0.744	0.9073	0.8183
6	2	H	34	29.2	17.61	8.665	2.694	0.7691	0.9675	0.8052
6	H	2	33	33.55	18.63	8.55	2.858	0.8184	0.9412	0.8601
6	H	H	34.93	31.16	18.67	8.744	2.8	0.8518	0.9681	0.8891
6	2	6	29.61	36.61	20.71	9.072	3.011	0.8883	0.9731	0.9039
6	6	2	30	37.9	19.82	8.467	3.05	0.8939	0.9407	0.9379
6	H	6	32.79	39.3	21.39	9.029	3.107	0.9082	0.9902	0.9165
6	6	H	33.09	35.9	20.42	8.955	3.009	0.9005	0.9697	0.9219
6	6	6	31.46	41.86	21.36	8.738	3.238	0.9319	0.9788	0.9449
		SED	3.253	2.84	0.6372	0.1449	0.0919	0.0203	0.0136	0.0183
		Residual Variance	63.86	48.66	2.45	0.127	0.0509	0.00249	0.00112	0.00204

Appendix 20. F_2 glasshouse summary of models fitted to generate predicted means

Tiller number:

Constant + *VRS1*+ *VRS3*+ *INT-C*

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	21.43	2	10.72	555	<0.001
VRS3	35.86	2	17.93	555	<0.001
INT-C	8.61	2	4.3	555	0.014

Central grain number:

Constant +*VRS1*+*VRS3*+*VRS1*.*VRS3*+*INT-C*

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	100.33	2	50.16	531	<0.001
VRS3	143.76	2	71.88	531	<0.001
VRS1.VRS3	19.25	4	4.81	531	<0.001
INT-C	14.98	2	7.49	531	<0.001

Central grain TGW:

Constant + *VRS1*+*VRS3*+*INT-C*+*VRS1*.*VRS3* + *VRS1*.*INT-C* + *VRS1*.*VRS3*. *INT-C*

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	227.98	2	113.99	515	<0.001
VRS3	106.15	2	53.08	515	<0.001
INT-C	28.64	2	14.32	515	<0.001
VRS1.VRS3	50.83	4	12.71	515	<0.001
VRS1.INT-C	23.95	4	5.99	515	<0.001
VRS1.VRS3. INT-C	32.33	12	2.69	515	0.002

Central grain area:

Constant + *VRS1+ VRS3+ INT-C+ VRS1.VRS3 + VRS1.INT-C + VRS1.VRS3. INT-C*

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	467.72	2	233.86	515	<0.001
VRS3	96.7	2	48.35	515	<0.001
INT-C	50.83	2	25.41	515	<0.001
VRS1.VRS3	83.66	4	20.92	515	<0.001
VRS1.INT-C	24.69	4	6.17	515	<0.001
VRS1.VRS3. INT-C	44.42	12	3.7	515	<0.001

Central grain length:

Constant + *VRS1+ VRS1. INT-C.VRS3*

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	49.17	2	24.58	515	<0.001
VRS1.INT-C.VRS3	47.04	24	1.96	515	0.004

Central grain width:

Constant+*VRS1+ INT-C+VRS3+VRS1.INT-C + VRS1.VRS3 + INT-C.VRS3+ VRS1. INT-C.VRS3*

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	349.1	2	174.55	515	<0.001
INT-C	36.64	2	18.32	515	<0.001
VRS3	81.06	2	40.53	515	<0.001
VRS1.INT-C	38.78	4	9.7	515	<0.001
VRS1.VRS3	80	4	20	515	<0.001
INT-C.VRS3	15.28	4	3.82	515	0.005
VRS. INT-C.VRS3	37.52	8	4.69	515	<0.001

Appendix 21. Homozygous row-type gene combinations field trial phenotypic means

a)Tiller number and whole plant grain phenotypic means

	Genotype *(VRS1,INT-C,VRS3)*	Tiller number	Whole plant TGW(g)	Whole plant grain area (mm^2)	Whole plant grain length(mm)	Whole plant grain width(mm)
	222	14.33	62.97	27.44	9.377	3.912
	226	12.79	47.89	22.32	8.4	3.51
	262	14.9	61.25	27.44	9.377	3.865
	266	9.8	46.18	22.32	8.4	3.463
	622	9.33	42.25	21.45	8.333	3.289
	626	11.4	43.57	22.06	8.602	3.361
	662	9.9	44.83	21.45	8.333	3.405
	666	8.4	46.15	22.06	8.602	3.478
	SED	1.09	1.108	0.446	0.1677	0.0342
Residual Variation	Entry	2.899	2.482	1.148	0.1852	0.002628
	row.col	7.218	8.837	1.029	0.104	0.00778

b) Central grain phenotypic means

	Genotype (*VRS1,INT-C,VRS3*)	Central Grain area(mm^2)	Central Grain Length (mm)	Central grain width(mm)	Central TGW(g)	Central grains per ear
	222	29.96	10.313	3.983	61.46	22.31
	226	27.65	9.861	3.805	55.8	18.44
	262	29.96	10.313	3.904	61.46	19.62
	266	27.65	9.861	3.824	55.8	15.75
	622	26.25	9.591	3.751	53.53	16.6
	626	26.78	9.847	3.714	52.45	16.37
	662	26.25	9.591	3.671	53.53	16.04
	666	26.78	9.847	3.734	52.45	15.82
	SED	0.4093	0.1513	0.03367	0.8886	0.7196
Residual Variation	Entry	1.067	0.16078	0.003732	3.503	0.267
	row.col	0.675	0.0657	0.00517	5.958	5.153

c) Lateral grain phenotypic means

	Genotype *(VRS1,INT-C,VRS3)*	Lateral grain number	Lateral grain area(mm^2)	Lateral grain length(mm)	Lateral grain width(mm)	Lateral TGW(g)	Lateral:central grain width ratio	Lateral:central grain length ratio	Lateral:central grain area ratio
	222								
	226	23.35	20.74	8.578	3.195	33.55	0.8328	0.8569	0.7296
	262								
	266	33.27	22.31	8.781	3.372	39.27	0.8851	0.8959	0.8166
	622	39.1	20.9	8.667	3.179	36.13	0.8464	0.8956	0.7794
	626	37.02	22.73	9.221	3.341	38.68	0.9027	0.9374	0.8495
	662	38.93	24.27	9.067	3.442	41.89	0.9387	0.9511	0.8969
	666	38.97	24.27	9.42	3.515	43.82	0.9377	0.956	0.9056
	SED	2.099	0.6695	0.2397	0.04534	1.596	0.01059	0.01174	0.01644
Residual Variation	Entry	12.25	1.5313	0.21045	0.004765	6.803	0.000207	0.000369	0.000762
	row.col	10.66	0.589	0.0512	0.00666	6.654	0.000458	0.000358	0.000634

Appendix 22. Homozygous row-type gene combinations field trial models fitted

Tiller number:

Fixed Model :Constant + *VRS1* + *VRS3* + *INT-C* + *VRS1*.*VRS3* + *VRS3*.*INT-C*

Random Model: row.col + Trial_entry

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	16.32	1	16.32	73.3	<0.001
VRS3	8.61	1	8.61	71.3	0.005
INT-C	11.28	1	11.28	73.3	0.001
VRS1.*VRS3*	7.77	1	7.77	72.4	0.007
VRS3.*INT-C*	7.77	1	7.77	72.1	0.007

Whole plant TGW

Fixed Model: Constant + *VRS1* + *VRS3* + *VRS1*.*VRS3* + *VRS1*.*INT-C*
Random Model: row.col + Trial_entry

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	227.15	1	227.15	73	<0.001
VRS3	147.71	1	147.71	73.3	<0.001
VRS1.*VRS3*	157.03	1	157.03	73.5	<0.001
VRS1.*INT-C*	12.32	2	6.16	73.5	0.003

Whole Plant Grain Area
Fixed Model: Constant + *VRS1* + *VRS3* + *VRS1*.*VRS3*
Random Model: row.col + Trial_entry

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	81.54	1	81.54	75.6	<0.001
VRS3	59.43	1	59.43	75	<0.001
VRS1.*VRS3*	81.32	1	81.32	75.1	<0.001

Whole Plant Grain Length

Fixed Model: Constant + *VRS1* + *VRS3* + *VRS1*.*VRS3*
Random Model: row.col+Trial_entry

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	7.11	1	7.11	74.5	0.009
VRS3	10.89	1	10.89	74	0.001
VRS1.*VRS3*	27.32	1	27.32	74.1	<0.001

Whole Plant Grain Width
Fixed Model: Constant + *VRS1* + *VRS3* + *VRS1*.*VRS3* + *VRS1*.*INT-C*
Random Model: row.col+Trial_entry

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	202.1	1	202.1	72	<0.001
VRS3	97.89	1	97.89	72.4	<0.001
VRS1.*VRS3*	134.29	1	134.29	72.5	<0.001
VRS1.*INT-C*	22.06	2	11.03	72.6	<0.001

Central Grain Per ear
Fixed: Constant + *VRS1* + *VRS3* + *INT-C* + *VRS1*.*VRS3* + *VRS1*.*INT-C*
Random Model: row.col+Trial_entry

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	35.91	1	35.91	69.3	<0.001
VRS3	34.98	1	34.98	70	<0.001
INT-C	23.04	1	23.04	70.5	<0.001
VRS1.*VRS3*	18.68	1	18.68	70.2	<0.001
VRS1.*INT-C*	6.62	1	6.62	70.1	0.012

Central Grain Area
Fixed Model: Constant + *VRS1* + *VRS3* + *VRS1*.*VRS3*

Random Model: row.col + Trial_entry

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	54.89	1	54.89	75.3	<0.001
VRS3	11.22	1	11.22	75	0.001
VRS1.*VRS3*	23.78	1	23.78	75	<0.001

Central Grain Length
Fixed Model: Constant + *VRS1* + *VRS1*.*VRS3*
Random Model: row.col + Trial_entry

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	7.64	1	7.64	74.3	0.007
VRS1.*VRS3*	12.04	2	6.02	74.2	0.004

Central Grain Width
Fixed Model: Constant + *VRS1* + *VRS3* + *VRS1*.*VRS3* + *VRS3*.*INT-C*

Random Model: row.col + Trial_entry

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	73.86	1	73.86	74.9	<0.001
VRS3	9.59	1	9.59	73.6	0.003
VRS1.*VRS3*	13.96	1	13.96	74.6	<0.001
VRS3.*INT-C*	7.05	2	3.53	74.4	0.034

Central TGW
Fixed Model: Constant + *VRS1* + *VRS3* + *VRS1*.*VRS3*
Random Model: row.col + Trial_entry

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
VRS1	88.1	1	88.1	75.4	<0.001
VRS3	30.97	1	30.97	74.9	<0.001
VRS1.*VRS3*	13.21	1	13.21	75	<0.001

Appendix 23. Expression analysis of row-type genes alignments of mutant and wild-type sequence

In all cases the transcription start sites are highlighted in pink; mutant polymorphisms are highlighted in yellow; nucleotides misaligned due to frameshift polymorphism are highlighted in red; premature transcription stop sites are highlighted in red, with the wild-type transcription stop site highlighted in turquoise. The nucleotides corresponding to the primers of qRTPCR amplicon are shown in green, with nucleotides targeted by the hydrolysis probe underlined where known.

INT-C

```
int-c.b3        ATGTTTCCTTTCTATGATTCCCCAAGCCCCATGGACTTACCCCTTTACCAGCAGCTGCAG 60
int-c.5         ATGTTTCCTTTCTATGATTCCCCAAGCCCCATGGACTTACCCCTTTACCAGCAGCTGCAG 60
                ************************************************************

int-c.b3        CTCAGCCCTCCCTCCCCAAAGGCGCCGGATCACCAATCCTTGCTCTACTACCATTCCTCC 120
int-c.5         CTCAGCCCTCCCTCCCCAAAGGCGCCGGATCACCAATCCTTGCTCTACTACCATTCCTCC 120
                ************************************************************

int-c.b3        CCTCCATTCGCCGCCGACCCTTTCCACCACAACTACCTCTGTGCCGGTGCCGGTGCCGGT 180
int-c.5         CCTCCATTCGCCGCCGACCCTTTCCACCACAACTACCTCTGTGCCGGTGCCGGTGCCGGT 180
                ************************************************************

int-c.b3        TCCGGTGCGGCCACGCCGCCAGCAGCCGAGATCGACGACCAGTCGCCGCCGGAGCTGCTT 240
int-c.5         TCCGGTGCGGCCACGCCGCCAGCAGCCGAGATCGACGACCAGTCGCCGCCGGAGCTGCTT 240
                ************************************************************

int-c.b3        CTGATGGATCAGGCGCCGGCGCCCAGGCCAGACGGGGTTGGAAAGGCACAAGGCCTGCAC 300
int-c.5         CTGATGGATCAGGCGCCGGCGCCCAGGCCAGACGGGGTTGGAAAGGCACAAGGCCTGCAC 300
                ************************************************************

int-c.b3        GGCGGTGGAGGCCTCGACAGCGCGGCCGCTAGGAAAGACCGTCACAGCAAGATATGCACC 360
int-c.5         GGCGGTGGAGGCCTCGACAGCGCGGCCGCTAGGAAAGACCGTCACAGCAAGATATGCACC 360
                ************************************************************

int-c.b3        GCCGGCGGGATGAGGGACCGGCGGATGCGGCTGTCCCTCGACGTCGCCCGCAAGTTCTTC 420
int-c.5         GCCGGCGGGATGAGGGACCGGCGGATGCGGCTGTCCCTCGACGTCGCCCGCAAGTTCTTC 420
                ************************************************************

int-c.b3        GCGCTCCAGGACATGCTCGGCTTCGACAAGGCCAGCAAGACGGTGCAATGGCTCCTCAAT 480
int-c.5         GCGCTCCAGGACATGCTCGGCTTCGACAAGGCCAGCAAGACGGTGCAATGGCTCCTCAAT 480
                ************************************************************

int-c.b3        ACGTCAAAGGGCGCCATCAAGGAGGTCATGACTGACGAGGCGTCCTCCGACTGCGAGGAG 540
int-c.5         ACGTCAAAGGGCGCCATCAAGGAGGTCATGACTGACGAGGCGTCCTCCGACTGCGAGGAG 540
                ************************************************************

int-c.b3        GACGGCTCCAGCAGCCTCTCCGTCGCCGACGGCAAGCACAAGCAGCCAGGGACGGAGGCT 600
int-c.5         GACGGCTCCAGCAGCCTCTCCGTCGCCGACGGCAAGCACAAGCAGCCAGGGACGGAGGCT 600
                ************************************************************

int-c.b3        GGAGGTGGTGATCACGCCGACGGGAAGAAGCCGGCGCCGAGGGCTTCCAAAAGGGCACCC 660
int-c.5         GGAGGTGGTGATCACGCCGACGGGAAGAAGCCGGCGCCGAGGGCTTCCAAAAGGGCACCC 660
                ************************************************************

int-c.b3        GCCAACCCCAAGCCGCAAAGGAAATTGGCCAGTGCGCACCTGATCCCCGACAAGGAGTCG 720
int-c.5         GCCAACCCCAAGCCGCAAAGGAAATTGGCCAGTGCGCACCTGATCCCCGACAAGGAGTCG 720
                ************************************************************

int-c.b3        AGGACAAAGGCGAGGGAGAGGGCAAGGGAGCGGACGAGGGAGAAGAACCGGATGCGATGG 780
int-c.5         AGGACAAAGGCGAGGGAGAGGGCAAGGGAGCGGACGAGGGAGAAGAACCGGATGCGATGG 780
                ************************************************************

int-c.b3        GTGACGCTCGCGTCCACAATCAACATCGAGCCGGCAACCACCGGCATGGCGGCGGCGAGG 840
int-c.5         GTGACGCTCGCGTCCACAATCAACATCGAGCCGGCAACCACCGGCATGGCGGCGGCGAGG 840
                ************************************************************

int-c.b3        CTGGACGAGTTGGTCACCAGCCCCAACAATTTGATCAATCGCTCCTCGTCCATGAACACG 900
int-c.5         CTGGACGAGTTGGTCACCAGCCCCAA-AATTTGATCAATCGCTCCTCGTCCATGAACACG 899
                ************************ ********************************

int-c.b3        CCAGGCGCTGAATTGGAGGAGGGGTGCTCGTCGTCCATGCCGAGCGAAGCGATCATGGCT 960
int-c.5         CCAGGCGCTGAATTGGAGGAGGGGTGCTCGTCGTCCATGCCGAGCGAAGCGATCATGGCT 959
```

```
                  ************************************************************

int-c.b3          GGCTTCGGCAATGGAGGGTACGGCAGCATCGGCAACTACTACCAGCACCAGCTGGAGCAG 1020
int-c.5           GGCTTCGGCAATGGAGGGTACGGCAGCATCGGCAACTACTACCAGCACCAGCTGGAGCAG 1019
                  ************************************************************

int-c.b3          CAATGGGAGCTCGGTGGAGTGGTGTTTGCCAACTCCCAGCACTACTGA 1068
int-c.5           CAATGGGAGCTCGGTGGAGTGGTGTTTGCCAACTCCCAGCACTACTGA 1067
                  ************************************************
```

VRS4

```
Vrs4              TGGCTCCTTGCTTTTCCATATTTTCTTGCCTACACACCTCAACTTTCCCTTGCTGCAGTC 60
vrs4.k            TGGCTCCTTGCTTTTCCATATTTTCTTGCCTACACACCTCAACTTTCCCTTGCTGCAGTC 60
                  ************************************************************

Vrs4              GTTGTAGGAGGAGAAGCTGTTCTCTATATCTCCCC-CCACTCCCACAGATCGATCACTGC 119
vrs4.k            GTTGTAGGAGGAGAAGCTGTTCTCTATATCTCCCCTCCACTCCCACAGATCGATCACTGC 120
                  ********************************** ************************

Vrs4              TGCCAGAGAGAGAATAGTTAGGCTGTCCTGCCTGATCATGTGCTAGCTCCTCCTCCTACA 179
vrs4.k            TGCCAGAGAGAGAATAGTTAGGCTGTCCTGCCTGATCATGTGCTAGCTCCCCCTCCTACA 180
                  ************************************************* *********

Vrs4              ATCCTTTGATTCACTGCCAAGGAGACAAGAACCACTCCTCCTGTTTGTTTGCCTGTTTTT 239
vrs4.k            ATCCTTTGATTCACTGCCAAGGAGACAAGAACCACTCCTCCTGTTTGTTTGCCTGTTTTT 240
                  ************************************************************

Vrs4              CCATTCGAAATTTCTCTCCTCCTTTGCCATGTTGTGAAGCTTGTGAATCTTGGAGGAAAC 299
vrs4.k            CCACTCGAAATTTCTCTCCTCCTTTGCCATGTTGTGAAGCTTGTGAATCTTGGAGGAAAC 300
                  *** ********************************************************

Vrs4              GGGCTGCAACAAGGAGAGGAAATTTCTCCCAAATATTTTCTTTGCCCATCTTCATCAGAC 359
vrs4.k            GGGCTGCAACAAGGAGAGGAAATTTCTCCCAAATATTTTCTTTGCCCATCTTCATCAGAC 360
                  ************************************************************

Vrs4              CATCAGATCTATATCCTCCTCTCCGGCCGCTAGTTGACCCCAAAGAGGTAAGGATGAGCA 419
vrs4.k            CATCAGATCTATATCCTCCTCTCCGGCCGCTAGTTGACCCCAAAGAGGTAAGGATGAGCA 420
                  ************************************************************

Vrs4              GGGTCTTTTGCTAGCTCGCTAGCTAGATGCACCTCAGAGAAATTTGAGAAGAAAGGAAAA 479
vrs4.k            GGGTCTTTTGCTAGCTCGCTAGCTAGATGCACCTCAGAGAAATTTGAGAAGAAAGGAAAA 480
                  ************************************************************

Vrs4              TTTTGCAGCCTGTCCTTCTCTTGCTACTATTATATTGTTCATGCTCTTCTTGAGATTTGA 539
vrs4.k            TTTTGCAGCCTGTCCTTCTCTTGCTACTATTATATTGTTCATGCTCTTCTTGAGATTTGA 540
                  ************************************************************

Vrs4              GCAAACATCATGCTATTTTTACTTGGAAATCTGTGCTTCTTGTGCCATATCATCATACAT 599
vrs4.k            GCAAACATCATGCTATTTTTACTTGGAAATCTGTGCTTCTTGTGCCATATCATCATACAT 600
                  ************************************************************

Vrs4              GCACGCATATGCTTTCTTACAGCTAGTGGATCAGGTTCTACATGCATTTGTTTTCATTGG 659
vrs4.k            GCACGCATATGCTTTCTTACAGCTAGTGGATCAGGTTCTACATGCATTTGTTTTCATTGG 660
                  ************************************************************

Vrs4              TTTTTTTTCAGCAGTCATATGTCAAATCATTCCCACTGCTATGAATGAATCTACAGTTTT 719
vrs4.k            TTTTTTTTCAGCAGTCATATGTCAAATCATTCCCACTGCTATGAATGAATCTACAGTTTT 720
                  ************************************************************

Vrs4              TCTGTGCGTCCTTGTAGTGAAGTTGAAAAAAGAAAGAAAAAGATACAGCTATTTCCCCCA 779
vrs4.k            TCTGTGCGTCCTTGTAGTGAAGTTGAAAAAAGAAAGAAAAAGATACAGCTATTTCCCCCA 780
                  ************************************************************

Vrs4              TTTCTTGCATCCAAAGCTCCCAAATATCAGATCAGAAACCCTAATCTCCTCCTTCAAC-- 837
vrs4.k            TTTCTTGCATCCAAAGCTCCCAAATATCAGATCAGAAACCCTAATCTCCTCCTTCAACCT 840
                  ************************************************************

Vrs4              --------CGATCCTCACTCCAAGAAAACCCTAACCCTAGAAGTAGTTAGCATTCGGAAG 889
vrs4.k            TCCCAGAACGATCCTCACTCCAAGAAAACCCTAACCCTAGAAGTAGTTAGCATTCGGAAG 900
```

```
                    **************************************************

Vrs4            AAACGAATGTTCTTTTCTCGGGAAAATGACTAACCATTTGACAGCTCTCTTAAATTCGCA 949
vrs4.k          AAACGAATGTTCTTTTCTCGGGAAAATGACTAACCATTTGACAGCTCTCTTAAATTCGCA 960
                ************************************************************

Vrs4            TATATATATATACGCAGGTCGCCGGCAATGGCATCCCCGTCGAGCACCGGCAACTCCATC 1009
vrs4.k          --TATATATATACGCAGGTCGCCGGCAATGGCATCCCCGTCGAGCACCGGCAACTCCATC 1018
                  ************************************************************

Vrs4            GTCTCCGTGGTGGTTGCAGCGGCCACGACACCGGGGGCCGGGGCGCCGTGCGCTGCGTGC 1069
vrs4.k          GTCTCCGTGGTGGTTGCAGCGGCCACGACACCGGGGGC-GGGGCGCCGTGCGCTGCGTGC 1077
                ************************************ ********************

Vrs4            AAGTTCCTACGGCGCAAGTGCCTCCCCGGCTGCGTGTTCGCGCCCTATTTCCCGCCGGAG 1129
vrs4.k          AAGTTCCTGCGGCGCAAGTGCCTCCCCGGCTGCGTGTTCGCGCCCTACTTCCCGCCGGAG 1137
                ********.*************************************** ************

Vrs4            GAGCCGCAGAAGTTCGCCAACGTGCACAAGGTGTTCGGCGCCAGCAACGTGACCAAGCTG 1189
vrs4.k          GAGCCGCAGAAGTTCGCCAACGTGCACAAGGTGTTCGGCGCCAGCAACGTGACCAAGCTG 1197
                ************************************************************

Vrs4            CTCAACGAGCTGCCGCCGCACCAGCGGGAAGACGCCGTGAGCTCGCTGGCCTACGAGGCG 1249
vrs4.k          CTCAACGAGCTGCCGCCGCACCAGCGGGAAGACGCCGTGAGCTCGCTGGCCTACGAGGCG 1257
                ************************************************************

Vrs4            GAGGCGCGGGTCAAGGACCCCGTCTACGGCTGCGTCGGCGCCATCTCCGTGCTCCAGCGC 1309
vrs4.k          GAGGCGCGGGTCAAGGACCCCGTCTACGGCTGCGTCGGCGCCATCTCCGTGCTCCAGCGC 1317
                ************************************************************

Vrs4            CAGGTCCACCGCCTCCAGAAGGAGCTCGACGCCGCGCACACCGAGCTCCTCCGGTACGCC 1369
vrs4.k          CAGGTCCACCGCCTCCAGAAGGAGCTCGACGCCGCGCACACCGAGCTCCTCCGGTACGCC 1377
                ************************************************************

Vrs4            TGCGGCGAGCTCGGCAGCATCCCCACCGCGCTCCCCGTTGTCACGGCCGGCGTCCCCAGC 1429
vrs4.k          TGCGGCGAGCTCGGCAGCATCCCCACCGCGCTCCCCGTTGTCACGGCCGGCGTCCCCAGC 1437
                ************************************************************

Vrs4            GGCAGGCTCTCATCCGCCGTAATGCCCTGCCCCGGCCAGCTCGCCGGCGGCATGTACAGC 1489
vrs4.k          GGCAGGCTCTCATCCGCCGTAATGCCCTGCCCCGGCCAGCTCGCCGGCGGCATGTACAGC 1497
                ************************************************************

Vrs4            GGTGGCGGTGGCGGTGGCTTCCGGAGGCTCGGGCTTGTGGACGCGATAGTACCGCAGCCC 1549
vrs4.k          GGTGGCGGTGGCGGTGGCTTCCGGAGGCTCGGGCTTGTGGACGCGATAGTACCACAGCCC 1557
                ************************************************.******

Vrs4            CCTCTTTCCGCCGGCTGCTACTACAATATGCGGAGCAACAACAACGCTGGAGGCAGCGTC 1609
vrs4.k          CCTCTTTCCGCCGGCTGCTACTACAATATGCGGAGCAACAACAACGCTGGAGGCAGCGTC 1617
                ************************************************************

Vrs4            GCTGCCGACGTGGCGCCCGTTCAGATCCCCTACGCCTCCATGGCGAATTGGGCCGTGAAC 1669
vrs4.k          GCTGCCGACGTGGCGCCCGTTCAGATCCCCTACGCCTCCATGGCGAATTGGGCCGTGAAC 1677
                ************************************************************

Vrs4            GCCATTAGCACCATTACCACCACCTCAGGATCAGAGAGCATTGGGATGGATCACAAGGAA 1729
vrs4.k          GCCATTAGCACCATTACCACCACCTCAGGATCAGAGAGCATTGGGATGGATCACAAGGAA 1737
                ************************************************************

Vrs4            GGAGGCGACAGCAGCATGTGAACTGACGCGCCCAACCAACAGTGAGTGGCCCGTTTGAGT 1789
vrs4.k          GGAGGCGACAGCAGCATGTGAACTGACGCGCCCAACCAACAGTGAGTGGCCCGTTTGAGT 1797
                ************************************************************

Vrs4            AAATCGTTGCACTGTAATCATGCGTGTATATATGCTAGCTTCATGAGTTATCTAATGCTG 1849
vrs4.k          AAATCGTTGCACTGTAATCATGCGTGTATATATGCTAGCTTCATGAGTTATCTAATGCTG 1857
                ************************************************************

Vrs4            CATTCTTGCCTTGTGCAGTGAATGTGGATTACCTCTCCTACTCTCAAGATTTATGGATGT 1909
vrs4.k          CATTCTTGCCTTGTGCAGTGAATGTGGATTACCTCTCCTACTCTCAAGATTTATGGATGT 1917
                ************************************************************

Vrs4            GGAGAATGATATGCGCTTTGGCCGGCCTTGACGCTCAAGGAAGGAAGCTTCGGGGAGACT 1969
```

```
vrs4.k          GGAGAATGATATGCGCTTTGGCCGGCCTTGACGCTCAAGGAAGGAAGCTTCGGGGAGACT 1977
                ************************************************************

Vrs4            GATGGGATCGAGCTAGCTCATTCCCAGGATAATTAAGCTAAGAAGCAATCTATCTATGAA 2029
vrs4.k          GATGGGATCGAGCTAGCTCATTCCCAGGATAATTAAGCTAAGAAGCAATCTATCTATGAA 2037
                ************************************************************

Vrs4            TCTATC 2035
vrs4.k          TCTATC 2043
                ******
```

VRS1

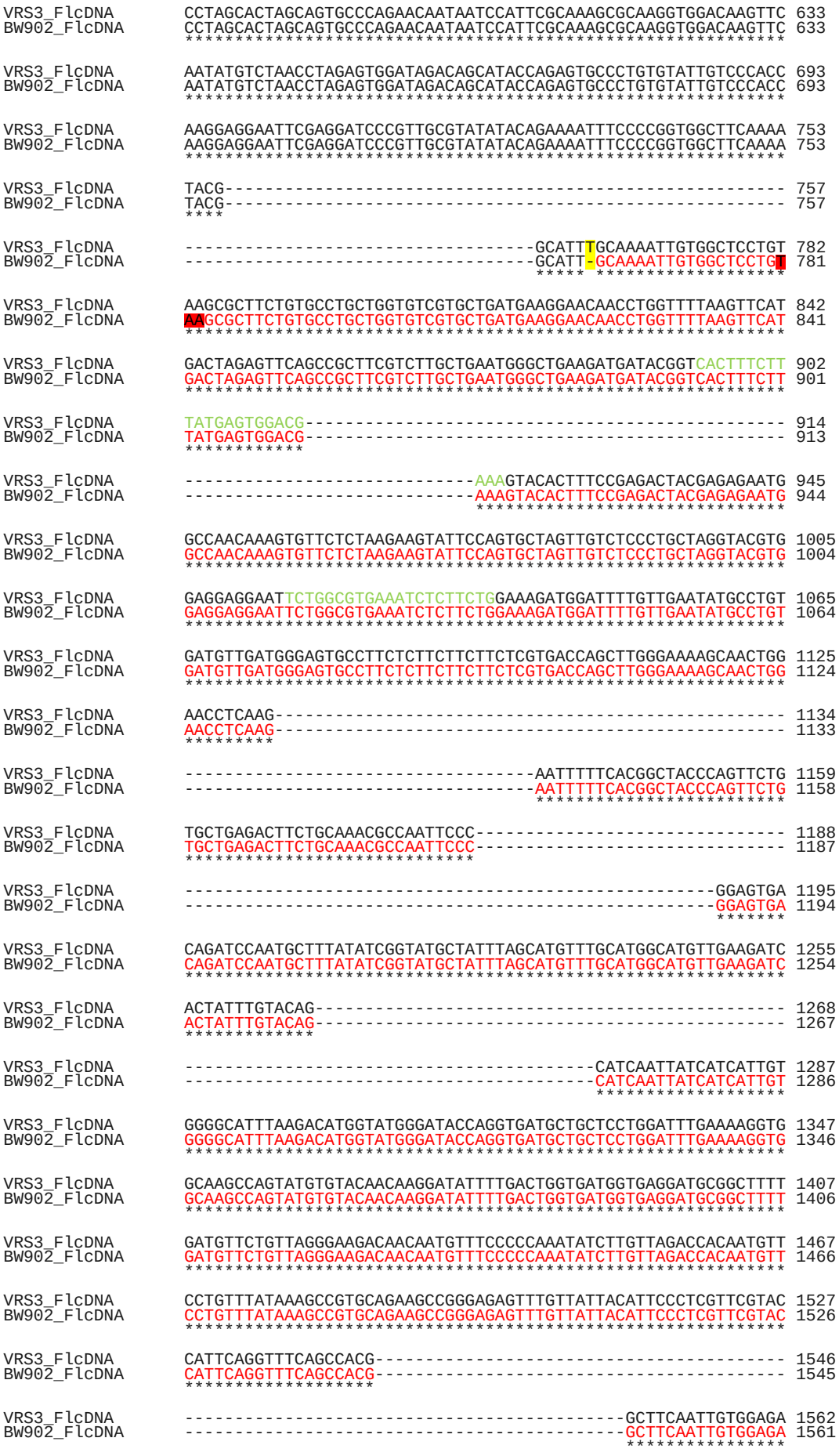
```
Vrs1.b3mRNA     ACGCACTCCCCTCCTTCAACTAGTGCTTTGCGGCCCGTGGTCCTCCTCTCGATCCAGTTC 60
vrs1.a1mRNA     ACGCACTCCCCTCCTTCAACTAGTGCTTTGCGGCCCGTGGTCCTCCTCTCGATCCAGTTC 60
                ************************************************************

Vrs1.b3mRNA     CTGAGCACACCAACAGGCAACAGAACAACCTACCGTGTCTCCCCTCCAATCTCCTCACGA 120
vrs1.a1mRNA     CTGAGCACACCAACAGGCAACAGAACAACCTACCGTGTCTCCCCTCCAATCTCCTCACGA 120
                ************************************************************

Vrs1.b3mRNA     TCCCTTCTTTCCCTCAGATCCGAACCGAAAGCATGGACAAGCATCAGCTCTTTGATTCAT 180
vrs1.a1mRNA     TCCCTTCTTTCCCTCAGATCCGAACCGAAAGCATGGACAAGCATCAGCTCTTTGGTTCAT 180
                ********************************************************* *****

Vrs1.b3mRNA     CCAACGTGGACACGACTTTCTTCGCGGCCAATGG-------------------------- 214
vrs1.a1mRNA     CCAACGTGGACACGACTTTCTTCGCGGCCAATGG-------------------------- 214
                ********************************

Vrs1.b3mRNA     ------------------------------------------------------------
vrs1.a1mRNA     ------------------------------------------------------------

Vrs1.b3mRNA     -------------------------------------CACGGCGCAGGGGGATACCAGC 236
vrs1.a1mRNA     -------------------------------------CACGGCGCAGGGGGAGACCAGC 236
                                                     ********************** ******

Vrs1.b3mRNA     AAGCAGAGGGCGCGGCGCAGGCGGCGGAGGTCGGCGAGGTGCGGCGGAGGGGATGGTGAC 296
vrs1.a1mRNA     AAGCAGAGGGCGCGGCGCAGGCGGCGGAGGTCGGCGAGGTGCGGCGGAGGGGATGGTGAC 296
                ************************************************************

Vrs1.b3mRNA     GGTGGGGAGATGGACGGAGGAGGGGACCCCAAGAAGCGGCGGCTCACCGACGAGCAGGCC 356
vrs1.a1mRNA     GGTGGGGAGATGGACGGAGGAGGGGACCCCAAGAAGCGGCGGCTCACCGACGAGCAGGCC 356
                ************************************************************

Vrs1.b3mRNA     GAGATTCTGGAGCTGAGCTTCCGGGAGGACCGCAAGCTGGAGACAGCCCGCAAGGTGTAT 416
vrs1.a1mRNA     GAGATTCTGGAGCTGAGCTTCCGGGAGGACCGCAAGCTGGAGACAGCCCGCAAGGTGTAT 416
                ************************************************************

Vrs1.b3mRNA     CTGGCCGCCGAGCTCGGGCTGGACCCCAAGCAGGTCGCCGTGTGGTTCCAGAACCGCCGC 476
vrs1.a1mRNA     CTGGCCGCCGAGCTCGGGCTGGACCCCAAGCAGGTCGCCGTGTGGTTCCAGAACCGCCGC 476
                ************************************************************

Vrs1.b3mRNA     GCGCGCCACAAGAACAAGACGCTCGAGGAGGAGTTCGCGAGGCTCAAGCACGCCCACGAC 536
vrs1.a1mRNA     GCGCGCCACAAGAACAAGACGCTCGAGGAGGAGTTCGCGAGGCTCAAGCACGCCCACGAC 536
                ************************************************************

Vrs1.b3mRNA     GCCGCCATCCTCCACAAATGCCACCTCGAGAACG-------------------------- 570
vrs1.a1mRNA     GCCGCCATCCTCCACAAATGCCACCTCGAGAACG-------------------------- 570
                ********************************

Vrs1.b3mRNA     ------------------------------------------------------------
vrs1.a1mRNA     ------------------------------------------------------------

Vrs1.b3mRNA     ---------------AGCTGCTGAGGCTGAAGGAGAGACTGGGAGCGACTGAGCAGGAGG 615
vrs1.a1mRNA     ---------------AGCTGCTGAGGCTGAAGGAGAGACTGGGAGCGACTGA-CAGGAGG 614
```

```
                                     ************************************ *******

Vrs1.b3mRNA      TGCGGCGCCTCAGGTCGGCAGCTGGGAGCCACGGGGCATCTGTGGATGGCGGACACGCCG 675
vrs1.a1mRNA      TGCGGCGCCTCAGGTCGGCAGCTGGGAGCCACGGGGCATCTGTGGATGGCGGACACGCCG 674
                 ************************************************************

Vrs1.b3mRNA      CTGGCGCCGTTGGCGTGTGCGGCGGGAGCCCGAGCTCGTCCTTCTCGACGGGAACCTGCC 735
vrs1.a1mRNA      CTGGCGCCGTTGGCGTGTGCGGCGGGAGCCCGAGCTCGTCCTTCTCGACGGGAACCTGCC 734
                 ************************************************************

Vrs1.b3mRNA      AGCAGCAGCCGGGTTTCAGCGGGGCAGACGTGCTGGGGCGGGACGATGACCTGATGATGT 795
vrs1.a1mRNA      AGCAGCAGCCGGGTTTCAGCGGGGCAGACGTGCTGGGGCGGGACGATGACCTGATGATGT 794
                 ************************************************************

Vrs1.b3mRNA      GCGTCCCCGAGTGGTTTTTAGCATGAATTAGAGTTTATGCTGGCTAAGCCGATAGCAGCG 855
vrs1.a1mRNA      GCGTCCCCGAGTGGTTTTTAGCATGAATTAGAGTTTATGCTGGCTAAGTCGATAGCAGCG 854
                 ********************************************** ***********

Vrs1.b3mRNA      TGGTCGAGTGTTTTTTAGCATGAAATCAGATCTCCATCTCCCATAAAATAGCCGAGATAG 915
vrs1.a1mRNA      TGGTCGAGTGTTTTTTAGCATGAAATCAGATCTCCATCTCCCATAAAATAGCCGAGATAG 914
                 ************************************************************

Vrs1.b3mRNA      CTGCTGCCGCCGCCAAATCCTCTATAGGGCTTCAAGATCGGCAGAAACCTCTAGAAATCA 975
vrs1.a1mRNA      CTGCTGCCGCCGCCAAATCCTCTATAGGGCTTCAAGATCGGCAGAAACCTCTAGAAATCA 974
                 ************************************************************

Vrs1.b3mRNA      TCTCCCCCCTCCGGAAAAGTCGCCTCTATTTGTCTCCATTGCCCGCGATGCAGCATCCGG 1035
vrs1.a1mRNA      TCTCCCCCCTCCGGAAAAGTCGCCTCTATTTGTCTCCATTGCCCGCGATGCAGCATCCGG 1034
                 ************************************************************

Vrs1.b3mRNA      TATAGCTGCTAAGACAGGCCGCCCCTAAATCGTTTCTCCAGCGATTTTAATCTTTGGTTT 1095
vrs1.a1mRNA      TATAGCTGCTAAGACAGGCCGCCCCTAAATCGTTTCTCCAGCGATTTTAATCTTTGGTTT 1094
                 ************************************************************

Vrs1.b3mRNA      TTAGCCTGTATATATGGGCTGTGATTTGAAGTTGAGACGAGCTGGACATCAACTGCACGC 1155
vrs1.a1mRNA      TTAGCCTGTATATATGGGCTGTGATTTGAAGTTGAGACGAGCTGGACATCAACTGCACGC 1154
                 ************************************************************

Vrs1.b3mRNA      TGATCGATTACTATTCTAGTTTGGCATAGTGTTAATTAAGTTTGGATGATCTCTAGGCGT 1215
vrs1.a1mRNA      TGATCGATTACTATTCTAGTTTGGCATAGTGTTAATTAAGTTTGGATGATCTCTAGGCGT 1214
                 ************************************************************

Vrs1.b3mRNA      GCGTTAAGTATGTAGATAGTGTTGATTAATGGCAAAAGCTTGCAAGTTAAGTGTAGTATT 1275
vrs1.a1mRNA      GCGTTAAGTATGTAGATAGTGTTGATTAATGGCAAAAGCTTGCAAGTTAAGTGTAGTATT 1274
                 ************************************************************

Vrs1.b3mRNA      GGCAGCTCTCTTGAAGATCAAATATGATGTGTGTTATC 1313
vrs1.a1mRNA      GGCAGCTCTCTTGAAGATCAAATATGATGTGTGTTATC 1312
                 ************************************
```

VRS3

```
VRS3_FlcDNA      GGGCTTCCCCATATCACCCTCCCCACCTCCCCCCGCCCCTCCCCCACAGCCAGAGACCCG 60
BW902_FlcDNA     GGGCTTCCCCATATCACCCTCCCCACCTCCCCCCGCCCCTCCCCCACAGCCAGAGACCCG 60
                 ************************************************************

VRS3_FlcDNA      TTCCGTCCCGTCCCGTTCCCTTCCTCTTCCGGCGACCCTCTCCCCCGCCGGCAGCCCCCG 120
BW902_FlcDNA     TTCCGTCCCGTCCCGTTTCCTTCCTCTTCCGGCGACCCGCTCCCCCGCCGGCAGCCCCCG 120
                 ********** ********** ********** **********

VRS3_FlcDNA      CCGTGGGGCCCCCATTGCTGGCTGGCCCTCAATTGCTTGCGCCAGCTCCCCCGCCCCCGC 180
BW902_FlcDNA     CCGTGGGGCCCCCATTGCTGGCTGGCCCTCAATTGCTTGCGCCAGCTCCCCCGCCCCCGC 180
                 ************************************************************

VRS3_FlcDNA      CTGCCTAGCGCGGGAAAATAACCGACGGGAGGAGCGAGGGCCGCGCCGCGCCTCGCCTTG 240
BW902_FlcDNA     CTGCCTAGCGCGGGAAAATAACCGACGGGAGGAGCGAGGGCCGCGCCGCGCCTCGCCTTG 240
                 ************************************************************

VRS3_FlcDNA      GCCCGGGAGGGATCCCCCACCGCCGCCCGCTCGCCCGCTCGCTCGTCAGGGGGGGAATGC 300
BW902_FlcDNA     GCCCGGGAGGGATCCCCCACCGCCGCCCGCTCGCCCGCTCGCTCGTCAGGGGGGGAATGC 300
                 ************************************************************

VRS3_FlcDNA      GGGACCCGCGCCGCCGGGGAATTGTCTAGACGGGCCTCCGCGACCGCGCCCCGCCCCGCC 360
```

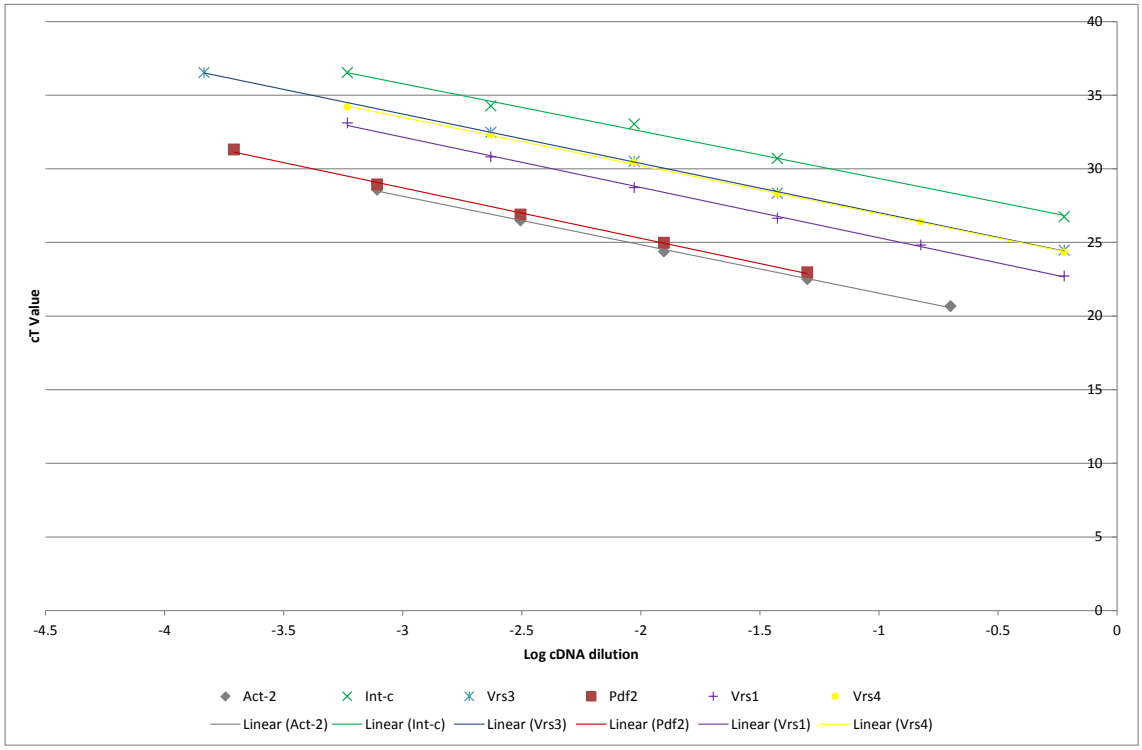
```
BW902_FlcDNA      GGGACCCGCGCCGCCGGGGAATTGTCTAGACGGGCCTCCGCGACCGCGCCCCGCCCCGCC 360
                  ************************************************************

VRS3_FlcDNA       GGAGTCATG---------------------------------------------------- 369
BW902_FlcDNA      GGAGTCATG---------------------------------------------------- 369
                  *********

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------GTGGAGGGAAGGAACTGTCTCCCT 393
BW902_FlcDNA      ------------------------------------GTGGAGGGAAGGAACTGTCTCCCT 393
                                                      ************************

VRS3_FlcDNA       GCAGAGGTCAGGAATGGGCTCGAGACACTGAAGAAGAGGAGACTTGAGAGGATGCGTTTG 453
BW902_FlcDNA      GCAGAGGTCAGGAATGGGCTCGAGACACTGAAGAAGAGGAGACTTGAGAGGATGCGTTTG 453
                  ************************************************************

VRS3_FlcDNA       AGTGCTCAGAAAGACGAAGGTGACAATCCTGCCATGGCTGCGAGGAGCGGCGGGGATTCA 513
BW902_FlcDNA      AGTGCTCAGAAAGACGAAGGTGACAATCCTGCCATGGCTGCGAGGAGCGGCGGGGATTCA 513
                  ************************************************************

VRS3_FlcDNA       CTGCGGACTCCAGCAAACTGTGGGGTCAGATTGCATGCCAACAATGTTACTGAAGGTTTA 573
BW902_FlcDNA      CTGCGGACTCCAGCAAACTGTGGGGTCAGATTGCATGCCAACAATGTTACTGAAGGTTTA 573
                  ************************************************************
```


VRS3_FlcDNA CCTAGCACTAGCAGTGCCCAGAACAATAATCCATTCGCAAAGCGCAAGGTGGACAAGTTC 633
BW902_FlcDNA CCTAGCACTAGCAGTGCCCAGAACAATAATCCATTCGCAAAGCGCAAGGTGGACAAGTTC 633
VRS3_FlcDNA AATATGTCTAACCTAGAGTGGATAGACAGCATACCAGAGTGCCCTGTGTATTGTCCCACC 693
BW902_FlcDNA AATATGTCTAACCTAGAGTGGATAGACAGCATACCAGAGTGCCCTGTGTATTGTCCCACC 693
VRS3_FlcDNA AAGGAGGAATTCGAGGATCCCGTTGCGTATATACAGAAAATTTCCCCGGTGGCTTCAAAA 753
BW902_FlcDNA AAGGAGGAATTCGAGGATCCCGTTGCGTATATACAGAAAATTTCCCCGGTGGCTTCAAAA 753
VRS3_FlcDNA TACG 757
BW902_FlcDNA TACG 757
VRS3_FlcDNA GCATTTGCAAAATTGTGGCTCCTGT 782
BW902_FlcDNA GCATT-GCAAAATTGTGGCTCCTGT 781
VRS3_FlcDNA AAGCGCTTCTGTGCCTGCTGGTGTCGTGCTGATGAAGGAACAACCTGGTTTTAAGTTCAT 842
BW902_FlcDNA AAGCGCTTCTGTGCCTGCTGGTGTCGTGCTGATGAAGGAACAACCTGGTTTTAAGTTCAT 841
VRS3_FlcDNA GACTAGAGTTCAGCCGCTTCGTCTTGCTGAATGGGCTGAAGATGATACGGTCACTTTCTT 902
BW902_FlcDNA GACTAGAGTTCAGCCGCTTCGTCTTGCTGAATGGGCTGAAGATGATACGGTCACTTTCTT 901
VRS3_FlcDNA TATGAGTGGACG 914
BW902_FlcDNA TATGAGTGGACG 913
VRS3_FlcDNA AAAGTACACTTTCCGAGACTACGAGAGAATG 945
BW902_FlcDNA AAAGTACACTTTCCGAGACTACGAGAGAATG 944
VRS3_FlcDNA GCCAACAAAGTGTTCTCTAAGAAGTATTCCAGTGCTAGTTGTCTCCCTGCTAGGTACGTG 1005
BW902_FlcDNA GCCAACAAAGTGTTCTCTAAGAAGTATTCCAGTGCTAGTTGTCTCCCTGCTAGGTACGTG 1004
VRS3_FlcDNA GAGGAGGAATTCTGGCGTGAAATCTCTTCTGGAAAGATGGATTTTGTTGAATATGCCTGT 1065
BW902_FlcDNA GAGGAGGAATTCTGGCGTGAAATCTCTTCTGGAAAGATGGATTTTGTTGAATATGCCTGT 1064
VRS3_FlcDNA GATGTTGATGGGAGTGCCTTCTCTTCTTCTTCTCGTGACCAGCTTGGGAAAAGCAACTGG 1125
BW902_FlcDNA GATGTTGATGGGAGTGCCTTCTCTTCTTCTTCTCGTGACCAGCTTGGGAAAAGCAACTGG 1124
VRS3_FlcDNA AACCTCAAG 1134
BW902_FlcDNA AACCTCAAG 1133
VRS3_FlcDNA AATTTTTCACGGCTACCCAGTTCTG 1159
BW902_FlcDNA AATTTTTCACGGCTACCCAGTTCTG 1158
VRS3_FlcDNA TGCTGAGACTTCTGCAAACGCCAATTCCC 1188
BW902_FlcDNA TGCTGAGACTTCTGCAAACGCCAATTCCC 1187
VRS3_FlcDNA GGAGTGA 1195
BW902_FlcDNA GGAGTGA 1194
VRS3_FlcDNA CAGATCCAATGCTTTATATCGGTATGCTATTTAGCATGTTTGCATGGCATGTTGAAGATC 1255
BW902_FlcDNA CAGATCCAATGCTTTATATCGGTATGCTATTTAGCATGTTTGCATGGCATGTTGAAGATC 1254
VRS3_FlcDNA ACTATTTGTACAG 1268
BW902_FlcDNA ACTATTTGTACAG 1267
VRS3_FlcDNA CATCAATTATCATCATTGT 1287
BW902_FlcDNA CATCAATTATCATCATTGT 1286
VRS3_FlcDNA GGGGCATTTAAGACATGGTATGGGATACCAGGTGATGCTGCTCCTGGATTTGAAAAGGTG 1347
BW902_FlcDNA GGGGCATTTAAGACATGGTATGGGATACCAGGTGATGCTGCTCCTGGATTTGAAAAGGTG 1346
VRS3_FlcDNA GCAAGCCAGTATGTGTACAACAAGGATATTTTGACTGGTGATGGTGAGGATGCGGCTTTT 1407
BW902_FlcDNA GCAAGCCAGTATGTGTACAACAAGGATATTTTGACTGGTGATGGTGAGGATGCGGCTTTT 1406
VRS3_FlcDNA GATGTTCTGTTAGGGAAGACAACAATGTTTCCCCCAAATATCTTGTTAGACCACAATGTT 1467
BW902_FlcDNA GATGTTCTGTTAGGGAAGACAACAATGTTTCCCCCAAATATCTTGTTAGACCACAATGTT 1466
VRS3_FlcDNA CCTGTTTATAAAGCCGTGCAGAAGCCGGGAGAGTTTGTTATTACATTCCCTCGTTCGTAC 1527
BW902_FlcDNA CCTGTTTATAAAGCCGTGCAGAAGCCGGGAGAGTTTGTTATTACATTCCCTCGTTCGTAC 1526
VRS3_FlcDNA CATTCAGGTTTCAGCCACG 1546
BW902_FlcDNA CATTCAGGTTTCAGCCACG 1545
VRS3_FlcDNA GCTTCAATTGTGGAGA 1562
BW902_FlcDNA GCTTCAATTGTGGAGA 1561
VRS3_FlcDNA GGCTGTCAATTTTGCTATTGGCGACTGGTTTCCTCTGGGCTCTCTGGCTAGCAAACGCTA 1622

```
BW902_FlcDNA      GGCTGTCAATTTTGCTATTGGCGACTGGTTTCCTCTGGGCTCTCTGGCTAGCAAACGCTA 1621
                  ************************************************************

VRS3_FlcDNA       CGCACTTCTGAACAGAACACCCTTTCTTGCGCACGAGGAGCTACTTTGTCTTTCTGCAAT 1682
BW902_FlcDNA      CGCACTTCTGAACAGAACACCCTTTCTTGCGCACGAGGAGCTACTTTGTCTTTCTGCAAT 1681
                  ************************************************************

VRS3_FlcDNA       GCTTCTTTCCCACAAACTGAGTGATCCAAAAACAATCAATTCTGAACATCCATACACTCA 1742
BW902_FlcDNA      GCTTCTTTCCCACAAACTGAGTGATCCAAAAACAATCAATTCTGAACATCCATACACTCA 1741
                  ************************************************************

VRS3_FlcDNA       ATATTGTGTGAAGTCTTCCTTTGTGAGGTTGATGCGAATGCAGCGGCGCACACGCAGCTT 1802
BW902_FlcDNA      ATATTGTGTGAAGTCTTCCTTTGTGAGGTTGATGCGAATGCAGCGGCGCACACGCAGCTT 1801
                  ************************************************************

VRS3_FlcDNA       ACTCGCTAAAATGGGTTCTCAGATATACTACAAGCCAAAAATGTATTCGAACCTATCCTG 1862
BW902_FlcDNA      ACTCGCTAAAATGGGTTCTCAGATATACTACAAGCCAAAAATGTATTCGAACCTATCCTG 1861
                  ************************************************************

VRS3_FlcDNA       TAGCATGTGCCGGCGTGATTGCTACGTTACACACGTGTCATGCGGATGCACCTTTGATCC 1922
BW902_FlcDNA      TAGCATGTGCCGGCGTGATTGCTACGTTACACACGTGTCATGCGGATGCACCTTTGATCC 1921
                  ************************************************************

VRS3_FlcDNA       TATCTGCCTTCATCATG---------------------------------------- 1939
BW902_FlcDNA      TATCTGCCTTCATCATG---------------------------------------- 1938
                  *****************

VRS3_FlcDNA       ----------------------------------------AACAAGAACTGCGGAGCTGCT 1960
BW902_FlcDNA      ----------------------------------------AACAAGAACTGCGGAGCTGCT 1959
                                                          **********************

VRS3_FlcDNA       CTTGTAAATCTGACCGGATTGTCTACGTCAGAGAAGACATACTGGAGTTAGAGGCTATAT 2020
BW902_FlcDNA      CTTGTAAATCTGACCGGATTGTCTACGTCAGAGAAGACATACTGGAGTTAGAGGCTATAT 2019
                  ************************************************************

VRS3_FlcDNA       ATAGAAAATTCGAGCAGGATATTCGCCTGGATAAGGAAACAAGTGCTAATATCTCGTATA 2080
BW902_FlcDNA      ATAGAAAATTCGAGCAGGATATTCGCCTGGATAAGGAAACAAGTGCTAATATCTCGTATA 2079
                  ************************************************************

VRS3_FlcDNA       AGCAAGCTGCGATTTCTGATATTGGTGTCGATCATGGTCCATCAGTTGGCACTGACCAGG 2140
BW902_FlcDNA      AGCAAGCTGCGATTTCTGATATTGGTGTCGATCATGGTCCATCAGTTGGCACTGACCAGG 2139
                  ************************************************************

VRS3_FlcDNA       ACATAAGCAACAGTGAAGCAAACTTGCTCGAAGCAAATGCTGCTGACTGTGGAAAGAGTT 2200
BW902_FlcDNA      ACATAAGCAACAGTGAAGCAAACTTGCTCGAAGCAAATGCTGCTGACTGTGGAAAGAGTT 2199
                  ************************************************************

VRS3_FlcDNA       CTCCTGCAACTTCATCGTTGACATCTTTTGCACTTCGTGACGGGTCTCTGCCTGCAGAAC 2260
BW902_FlcDNA      CTCCTGCAACTTCATCGTTGACATCTTTTGCACTTCGTGACGGGTCTCTGCCTGCAGAAC 2259
                  ************************************************************

VRS3_FlcDNA       CAAAG---------------------------------------------------- 2265
BW902_FlcDNA      CAAAG---------------------------------------------------- 2264
                  *****

VRS3_FlcDNA       -------------------------GTCCATGCAGCTCGAACCGACCAAATTTGGTCAA 2299
BW902_FlcDNA      -------------------------GTCCATGCAGCTCGAACCGACCAAATTTGGTCAA 2298
                                          **********************************

VRS3_FlcDNA       TTGCTAAGCAGGCCATAAAAACATCATCAGTGGAAGGAAATGGTGCGCTGGATGGCAATT 2359
BW902_FlcDNA      TTGCTAAGCAGGCCATAAAAACATCATCAGTGGAAGGAAATGGTGCGCTGGATGGCAATT 2358
                  ************************************************************

VRS3_FlcDNA       CATCCTGCATGGCTGATGCTTGCAATGAAATTAGTTCATGTAATGCTTCACCCATGGAAT 2419
BW902_FlcDNA      CATCCTGCATGGCTGATGCTTGCAATGAAATTAGTTCATGTAATGCTTCACCCATGGAAT 2418
                  ************************************************************

VRS3_FlcDNA       ATAGTGGAAATTCTGATTCTGATTCTGAAATCTTCCGAGTCAAGCGCAGATCCAGCATAT 2479
BW902_FlcDNA      ATAGTGGAAATTCTGATTCTGATTCTGAAATCTTCCGAGTCAAGCGCAGATCCAGCATAT 2478
                  ************************************************************

VRS3_FlcDNA       TAGGAAGATCTGCTCCTGACACAAAGACAACAAACTTATCTGAACAGAAG---------- 2529
BW902_FlcDNA      TAGGAAGATCTGCTCCTGACACAAAGACAACAAACTTATCTGAACAGAAG---------- 2528
                  ************************************************

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ------------------------------------------------------------
BW902_FlcDNA      ------------------------------------------------------------

VRS3_FlcDNA       ----------------------------------------------------------GTTC 2533
BW902_FlcDNA      ----------------------------------------------------------GTTC 2532
                                                                           ****

VRS3_FlcDNA       TGAAGCGGCTGAAGGAGGCATCCCCAGAAACACAACATGAGAATAAGCGACATGAAGAAG 2593
BW902_FlcDNA      TGAAGCGGCTGAAGGAGGCATCCCCAGAAACACAACATGAGAATAAGCGACATGAAGAAG 2592
                  ************** ************************************************

VRS3_FlcDNA       ACTCTGAGCGGACTTCAGTTCCCTCAGTTAGTAGGAGGCACAATAAGTCAAATTCTGGCT 2653
BW902_FlcDNA      ACTCTGAGCGGACTTCAGTTCCCTCAGTTAGTAGGAGGCACAATAAGTCAAATTCTGGCT 2652
                  ************************************************************
```

```
VRS3_FlcDNA       CTTCTGAGGAAGATAGAGAGGACATGGTTCCCATTGCCTGGAGGATGAAGCGGCGGCAGC 2713
BW902_FlcDNA      CTTCTGAGGAAGATAGAGAGGACATGGTTCCCATTGCCTGGAGGATGAAGCGGCGGCAGC 2712
                  ************************************************************

VRS3_FlcDNA       TGGAAGCTCAACAAGGCGACACCAGTTATGCCGCGCTGCAGTCGAAGGCGTATCCATCTA 2773
BW902_FlcDNA      TGGAAGCTCAACAAGGCGACACCAGTTATGCCGCGCTGCAGTCGAAGGCGTATCCATCTA 2772
                  ************************************************************

VRS3_FlcDNA       CCGGCAGCTGTTCCCGGCAGCAGTTTGCGGAGGCAACTAAAGACGCAGCCTCAGAGGTCC 2833
BW902_FlcDNA      CCGGCAGCTGTTCCCGGCAGCAGTTTGCGGAGGCAACTAAAGACGCAGCCTCAGAGGTCC 2832
                  ************************************************************

VRS3_FlcDNA       GGCCAAAGCGGGTGAAGATCCGGTTGCCTCGTAGCGCTGCAAACAGGCTGGTTGAGGGGC 2893
BW902_FlcDNA      GGCCAAAGCGGGTGAAGATCCGGTTGCCTCGTAGCGCTGCAAACAGGCTGGTTGAGGGGC 2892
                  ************************************************************

VRS3_FlcDNA       GGCAGCAGCAGGTCAGTTCAGGCCAGGGATTTGCAGTGGATGACAAGCCACCTGGGTTTT 2953
BW902_FlcDNA      GGCAGCAGCAGGTCAGTTCAGGCCAGGGATTTGCAGTGGATGACAAGCCACCTGGGTTTT 2952
                  ************************************************************

VRS3_FlcDNA       GGCATACGGTTTAGGGTAGCCGCCCTCAGGTGCTAACAAGCTTGACAGGTAGGACTCGAT 3013
BW902_FlcDNA      GGCATACGGTTTAGGGTAGCCGCCCTCAGGTGCTAACAAGCTTGACAGGTAGGACTCGAT 3012
                  ************************************************************

VRS3_FlcDNA       TCTTTGTTTAGGTGCATTAACTACACGAAGGCGATTTCATTTTGGTTAGCACAGATGCTG 3073
BW902_FlcDNA      TCTTTGTTTAGGTGCATTAACTACACGAAGGCGATTTCATTTTGGTTAGCACAGATGCTG 3072
                  ************************************************************

VRS3_FlcDNA       ACTGACTTGGGTTTAGAAAGGAGAGAGGAGGGTAGGGCTGATTGCTTGGTTAATCCATTC 3133
BW902_FlcDNA      ACTGACTTGGGTTTAGAAAGGAGAGAGGAGGGTAGGGCTGATTGCTTGGTTAATCCATTC 3132
                  ************************************************************

VRS3_FlcDNA       TTTGTGCAGCAGAGCCTCCAATTTGAGGCCGGCCTAGGTAGGCACACGGTTATTAGAAGA 3193
BW902_FlcDNA      TTTGTGCAGCAGAGCCTCCAATTTGAGGCCGGCCTAGGTAGGCACACGGTTATTAGAAGA 3192
                  ************************************************************

VRS3_FlcDNA       CAAGACAATTGTTCATCACGTTGTTTAGAGGCTTCCTCTGAGTCATCGCTTGCTGGCTGC 3253
BW902_FlcDNA      CAAGACAATTGTTCATCACGTTGTTTAGAGGCTTCCTCTGAGTCATCGCTTGCTGGCTGC 3252
                  ************************************************************

VRS3_FlcDNA       TCTTTGATGCCTGCATTGCCCGTGACGGAATTGGGTCTCCATTCTGCAAGTGTTTCTTTT 3313
BW902_FlcDNA      TCTTTGATGCCTGCATTGCCCGTGACGGAATTGGGTCTCCATTCTGCAAGTGTTTCTTTT 3312
                  ************************************************************

VRS3_FlcDNA       GCTTAGCACAGCTGTATATCTCTTTCTTGTTTTTATCATCGAGAACAAGTGGAGCGTTAG 3373
BW902_FlcDNA      GCTTAGCACAGCTGTATATCTCTTTCTTGTTTTTATCATCGAGAACAAGTGGAGCGTTAG 3372
                  ************************************************************

VRS3_FlcDNA       CTGTTGCTAACTGTATTCTTTCGTACACCTGCTCACGAAATATTCAATCGATCAATCATT 3433
BW902_FlcDNA      CTGTTGCTAACTGTATTCTTTCGTACACCTGCTCACGAAATATTCAATCGATCAATCATT 3432
                  ************************************************************

VRS3_FlcDNA       CTTTTCGG 3441
BW902_FlcDNA      CTTTTCGG 3440
                  ********
```

Appendix 24. qRTPCR primer standard curves and efficiencies



Gene target	Standard Curve	R^2	slope	Efficiency (10^-1/slope)	% Primer Efficiency
Act2	y = -3.2877x + 18.267	0.999	-3.2877	2.014484848	101.45
Int-c	y = -3.2183x + 26.116	0.9954	-3.2183	2.045140045	104.51
Vrs3	y = -3.3467x + 23.676	0.9998	-3.3467	1.98976508	98.98
Pdf2	y = -3.4274x + 18.428	0.9983	-3.2093	2.04924757	104.92
Vrs1	y = -3.4174x + 21.898	0.9989	-3.4174	1.961643667	96.16
Vrs4	y = -3.2768x + 23.661	0.9996	-3.2768	2.019183465	101.92

Appendix 25. qRTPCR NRQ ANOVA means

Gene Target	Growth Stage	Mean Square Genotype	Mean Square Rep	Residual Mean Square	Genotype Probability	SED	Log Transformed Means					
							Bowman	BW898(vrs1.a)	BW901(vrs2.e)	BW902(vrs3.f)	BW903(vrs4.k)	BW421(int-c.5)
Vrs1	1	10.541	2.936	1.145	0.002	0.874	0	-0.100	-0.920	-1.300	-5.020	-0.600
Vrs1	2	5.71168	0.3084	0.05228	<.001	0.187	0	0.131	0.112	-1.014	-3.453	-0.486
Vrs1	3	3.97422	0.07901	0.0401	<0.001	0.164	0	0.721	0.719	-0.749	-2.349	-0.069
Vrs3	1	0.5283	1.4713	0.1734	0.063	0.34	0	0.044	-0.270	-1.042	0.014	-0.047
Vrs3	2	0.2151	1.0007	0.2006	0.431	0.366	0	0.070	0.110	-0.390	0.430	-0.090
Vrs3	3	0.36827	0.33124	0.09384	0.036	0.25	0	0.438	0.716	-0.191	0.573	0.170
Vrs4	1	0.20006	0.07235	0.04457	0.021	0.172	0	0.278	-0.041	-0.294	-0.454	0.026
Vrs4	2	0.3723	0.4336	0.1232	0.064	0.287	0	0.004	0.086	-0.802	-0.525	-0.224
Vrs4	3	0.08291	0.78581	0.05418	0.273	0.19	0	0.299	-0.048	0.083	-0.164	-0.114
Int-C	1	1.4857	0.3418	0.3482	0.024	0.482	0	0.220	-0.020	-1.170	-1.440	-0.060
Int-C	2	1.33	0.2194	0.1301	0.001	0.295	0	-0.002	-0.513	-1.161	-1.585	-0.139
Int-C	3	0.9012	2.7676	0.2305	0.037	0.392	0	0.360	-0.580	-0.310	-1.250	-0.380